

BRUCE A. CHADWICK
HOWARD M. BAHR
STAN L. ALBRECHT

SOCIAL
SCIENCE
RESEARCH
METHODS



Digitized by the Internet Archive
in 2010

Social Science Research Methods

Bruce A. Chadwick

Howard M. Bahr

Stan L. Albrecht

all Brigham Young University

Library of Congress Cataloging in Publication Data

CHADWICK, BRUCE A.

Social science research methods.

Bibliography

Includes index.

1. Social sciences—Research—Methodology.
2. Social surveys. I. Bahr, Howard M. II. Albrecht, Stan L. III. Title.

H62.C3473 1984 300'.72 83-19049

ISBN 0-13-818336-8

Editorial/production supervision
and interior design: Barbara Kelly Kittle
Cover design: Ben Santora
Manufacturing buyer: John Hall

© 1984 by Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632

*All rights reserved. No part of this book may be
reproduced, in any form or by any means,
without permission in writing from the publisher.*

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

ISBN 0-13-818336-8

Prentice-Hall International, Inc., London
Prentice-Hall of Australia Pty. Limited, Sydney
Editora Prentice-Hall do Brasil, Ltda., Rio de Janeiro
Prentice-Hall Canada Inc., Toronto
Prentice-Hall of India Private Limited, New Delhi
Prentice-Hall of Japan, Inc., Tokyo
Prentice-Hall of Southeast Asia Pte. Ltd., Singapore
Whitehall Books Limited, Wellington, New Zealand

contents

Preface xi

1 KNOWLEDGE, SCIENCE, AND RESEARCH 1

Introduction 2
Knowledge 3
Scientific Method 6
Limits of Scientific Observation 10
Research And Theory 14
Ethics of Doing Research with Human Subjects 15
Why Do Research? 20
Summary 22
References 23

2 RESEARCH DESIGN 25

Introduction 26
Selecting Problems to Research 28

Types of Research Design	32
Stages in Research Design: Preparing a Plan	34
Principles of Effective Design	37
Research Language	42
Summary	48
References	49

3 SAMPLING 51

Introduction	52
Random Sampling	53
Nonrandom Sampling	65
Sample Size	67
Summary	69
References	70

4 OBSERVATIONAL METHODS 72

Introduction	73
Advantages of Observational Methods	74
Problems with Observations	76
Doing Observational Research	78
Analyzing Observational Data	87
Illustrative Examples	88
Summary	96
References	97

5 SURVEY RESEARCH: INTERVIEW STUDIES 99

Introduction	100
Advantages of Survey Research	101
Disadvantages of Surveys	102
The Interview	102
Creating an Interview Schedule	115
Pretesting the Interview Schedule	120
Conducting the Interview	120
Other Variations: The Telephone Interview	126

Collecting Data by Interview: An Illustrative Example	128
Summary	130
References	132

6 SURVEY RESEARCH QUESTIONNAIRE STUDIES 134

Introduction	135
Advantages of Survey by Questionnaire	137
Disadvantages of Survey by Questionnaire	138
Doing a Survey by Questionnaire	139
An Illustrative Example: Equality of Educational Opportunity	153
Summary	160
References	162

7 EXPERIMENTAL RESEARCH 163

Introduction	164
Types of Experiments	166
Strengths and Weaknesses of Experimental Research	167
Experimental Subjects	171
Steps of Experimental Research	174
Illustrative Experiments	187
Summary	202
References	203

8 QUALITATIVE RESEARCH 205

Introduction	206
Strengths of Qualitative Research	211
Problems of Qualitative Research	212
Design of Qualitative Research	215
Doing Qualitative Research	218
Illustrative Examples	225
Summary	235
References	236

9 CONTENT ANALYSIS 238

- Introduction 239
- Strengths and Weaknesses of Content Analysis 243
- Sources of Material 245
- Computer Assisted Content Analysis 247
- Doing Content Analysis 248
- Exemplary Content Analysis Studies 250
- Summary 255
- References 256

10 SECONDARY ANALYSIS 258

- Introduction 259
- Advantages and Disadvantages of Secondary Analysis 261
- Doing Secondary Analysis 267
- Illustrative Examples 271
- Summary 276
- References 279

11 EVALUATION RESEARCH 281—coauthor, Douglas Hooper

- Introduction 282
- The Organizational Context of Evaluation Research 285
- Evaluation Design 290
- Evaluation Methods 296
- An Example: The Juvenile Diversion Problem Evaluation 304
- Summary 309
- References 310

12 SOCIAL IMPACT ASSESSMENT 313

- Introduction 314
- Why Do Social Impact Assessment? 316
- What Is Included in a Social Impact Assessment? 320
- Doing Social Impact Assessments 322
- Impact Mitigation And Monitoring 339

Summary	341
References	342

13 ANALYZING DATA 344

Introduction	345
Coding	346
Data Entry	356
Descriptive Analysis	358
Cross-tabulation and Contingency Tables	367
Testing Relationships between Variables	369
Indexing and Scaling	384
Summary	387
References	388

14 WRITING ABOUT RESEARCH 390

Introduction	391
The Information Explosion	392
In Defense of Social Science Writing	393
Writing as Scientific Creativity	397
Three Basic Issues in Writing	399
The Process of Writing	402
Summary	405
References	405

15 THE UTILIZATION OF SOCIAL SCIENCE RESEARCH 408

Introduction	409
Some Uses of Social Science Research	411
Social Science Research and National Policy Making	412
Barriers to Social Research Utilization	416
Social Research and the Individual Consumer	423
Summary	429
References	430

Glossary 433

Index 445

preface

We have prepared this text for students taking their first class in research methods. We have written from the assumption that knowledge in any scientific discipline is advanced through the careful and competent application of a variety of techniques and procedures, "tools of the trade." This book is designed to acquaint the student with the tools of social science.

A broad range of research methods is presented. Most social science disciplines use several of the methods discussed and some disciplines use them all. However, there is a tendency for each social science discipline to emphasize some methods and ignore others. Many sociologists and political scientists collect their data through the use of questionnaires and surveys, although others use participant observation or secondary data analysis. Psychologists tend to emphasize experimental design. Anthropologists favor formal observation and participant observation; although many contemporary cultural anthropologists also do a lot of interviewing. Historians and economists emphasize secondary analysis and content analysis. We emphasize throughout the book that, regardless of a scientist's discipline, one of the best ways to improve the quality of research is to use two or more different research methods to study a particular problem. We feel that learning about each of these methods will make the reader a

better consumer, as well as doer, of research, whatever his or her academic major.

While research aimed at testing scientific theories and advancing scientific knowledge is stressed, we also consider at length social research done in applied settings to assist political, business, community, and other leaders to build and manage successful social programs. The chapters on evaluation research and social impact assessment, for example, explain how social research is used to determine the effectiveness of ongoing social programs and to assess the potential impacts of large-scale projects like the construction of nuclear power plants, the building of new highway systems, and so on. Other chapters describe and illustrate both "pure" and "applied" research.

Collecting data via the application of one or more research methods is not the final step in the research process. Therefore, we have prepared chapters on data analysis procedures and research report writing. Good research has little impact unless it is made public, and even then it may be of little consequence unless the analysis is competently done and the results written in an interesting, informative manner. We have also included a final chapter on how social research has been and might be used to improve communities and societies generally. It is our conviction that the quality of the social context can be improved and that the negative consequences of "social problems" can be ameliorated by the effective use of the findings of competently performed social research.

We recognize that many, probably most, students taking a course in research methods will not become professional researchers. Accordingly, we have tried to give special emphasis to the task everyone faces regularly, that of *consuming* and *interpreting* research that is done by others. A course in research methods should help students become better able to recognize good research and be more aware of how such research can be applied to real problems. The information regularly reported in the media, provided by public opinion polls, community surveys, or social experiments—research concerned with such important topics as inflation, legalized abortion, capital punishment, and pornography—can be more accurately assessed if the "consumer" has a basic understanding of research methods.

We have included as many specific examples of research as space would allow to illustrate the various research strategies. Several "classics" are reviewed because they represent creative innovations, successful applications of specific techniques, or important research failures, all of which are valuable in increasing our understanding of the research process. In addition, as often as possible we have used studies that have had important impacts on our society to demonstrate that research can "make a difference" in the political process and in the decisions people make about what is and what may be. The examples are not presented as the definitive ways to conduct research but, rather, as illustrations of how particular important

research questions have been studied. We have tried to provide examples that highlight the role of creativity in the research process. Throughout the text we try to make the point that there are no hard, fast rules or formulas that tell the researcher or the student exactly what to do in a given situation. Rather, it is acknowledged that careful thought and creativity are essential ingredients in any successful social research project, as they are in any human enterprise.

Most importantly we hope that students will glimpse the excitement of doing research. Testing theory and trying to understand and solve social problems may be serious business, but the process of research itself is enjoyable and exciting.

We express appreciation to the editors at Prentice-Hall, Ed Stanford and Susan Taylor, for their encouragement and support. Barbara Kelly Kittle deserves special thanks for serving as production editor and shepherding the manuscript to publication. Special appreciation is extended to Douglas Hooper, who co-authored Chapter II, *Evaluation Research*. The following have served as reviewers of various portions of the manuscript and their assistance and suggestions are greatly appreciated: Professor Thomas Dietz, George Washington University; Professor James R. Marshall, SUNY Buffalo; and Professor Theodore C. Wagenaar, Miami University, Ohio.

Ruth Barlow, Barbara Jenkins, Lori Vernon, and Mie Walker each went beyond the call of duty in typing, copying, cutting, and pasting the manuscript in order to meet deadlines, and we are grateful for their support.

Knowledge, Science, and Research

one

I. Introduction

II. Knowledge

Types of knowledge

Avenues to knowledge

III. Scientific Method

The assumptions of science

Characteristics of research

IV. Limits of Scientific Observation

*You cannot observe something
without changing it*

*You cannot observe something
without misperceiving it*

*You cannot interpret (attribute
meaning to) an observation
without misrepresenting it*

*You cannot communicate an
interpretation of an obser-
vation without an additional
misrepresentation*

V. Research and Theory

VI. Ethics of Doing Social Research with Human Subjects

VII. Why Do Research?

VIII. Summary

I. INTRODUCTION

This is a book about doing social science research. We begin with the assumption that knowledge about our social world accumulates, in significant part, from careful research about that world. In the chapters that follow we will describe in some detail the variety of methods that social scientists use in accumulating knowledge. In the social sciences, as in any scientific discipline, knowledge is advanced through the careful collection, proper analysis, and competent interpretation of research data. However, as Theodore Newcomb (1966: 1) noted many years ago, no research findings are better than the methods used to obtain them.

The repertoire of methods available for social science research has both expanded and become refined over the years. The range of tools available is broad and diverse, and so the task of understanding them and their application is challenging.

The principal purpose of this book is to teach the student how to do research, but it has another goal of almost equal importance. We recognize that comparatively few students taking their first course in research methodology will devote a major portion of their lives to doing social research. Most of them will follow other careers. Nevertheless, all students who take such a course will be daily "consumers" of research, as we all are. For example, when we consider decisions about important purchases, when

we evaluate the relative merits of various political programs, when we make decisions about investments, and so on, we frequently take into account *research* that has been done on these topics. Thus, we actively use research even though we may not produce it ourselves. Therefore, some knowledge about how research is done is of value to us all, whatever the nature of our future employment or activity.

Before turning to a discussion of specific research methods, we will review some basic principles of the accumulation of scientific knowledge. We will also identify some of the important problems that are associated with doing research on human subjects as opposed to nonthinking, non-reactive, inanimate objects or substances. As we shall see, the very nature of the subject matter of social science research creates important challenges, but it also increases the excitement and satisfaction that comes from research that is done well. Probably the single most important issue in conducting social research is protecting human subjects from being harmed, and we will discuss this responsibility in detail.

II. KNOWLEDGE

It is customary for teachers of introductory courses in the social sciences to explain that *science*, whether physical or social, is not concerned with "truth." That term connotes an absolute, metaphysical, unchangeable reality that is, by definition, beyond the reach of scientific procedure. On the other hand, words like "findings," "observations," and "results" apply to the world of appearances, which is the world accessible to the scientist.

Social science helps us to "understand" ourselves and our social world, but understanding is, again by definition, not necessarily based on reality or truth. Although some social scientists would be uncomfortable in defining their role as seekers of knowledge, that definition is, in our view, the most appropriate one. Knowledge refers to an acquaintance with facts, especially bodies of facts that are organized into principles of human behavior. The term "knowledge" is a stronger word than "understanding" and better suited to describe the end product of research. In this book we are concerned with methods that help people to accumulate knowledge.

Types of Knowledge

There are many ways of categorizing human knowledge. From the standpoint of social inquiry, perhaps the most relevant way is with reference to how it is obtained. Henle (1969: 11–12) lists five distinct ways of knowing: the humanistic, the scientific, the philosophical, the mathematical, and the theological. He also says that "these five ways of knowing

are irreducibly different, that they therefore give rise to formally different groups of discipline."

Boulding (1978: 171–172) suggests there are three kinds of human knowledge: folk knowledge, literary knowledge, and scientific knowledge. Folk knowledge is the knowledge acquired from day-to-day experience; it may or may not be based on some kind of testing, either in personal experience or in the experience of respected others. More abstract is the "imaginary" or literary knowledge, which is not subject to the normal checks and tests that folk knowledge is but which survives because of symbolic reality; literary knowledge is created in an abstracting process whereby essential realities are distilled from human experience and used to illustrate potential human capabilities. The third kind of knowledge is scientific knowledge, which "has achieved its extraordinary success by combining the testing which is characteristic of folk knowledge with the 'theorizing' which is characteristic of literary knowledge." The major activities of **empirical** science, says Boulding, are description and testing, and the result is the expansion of knowledge.

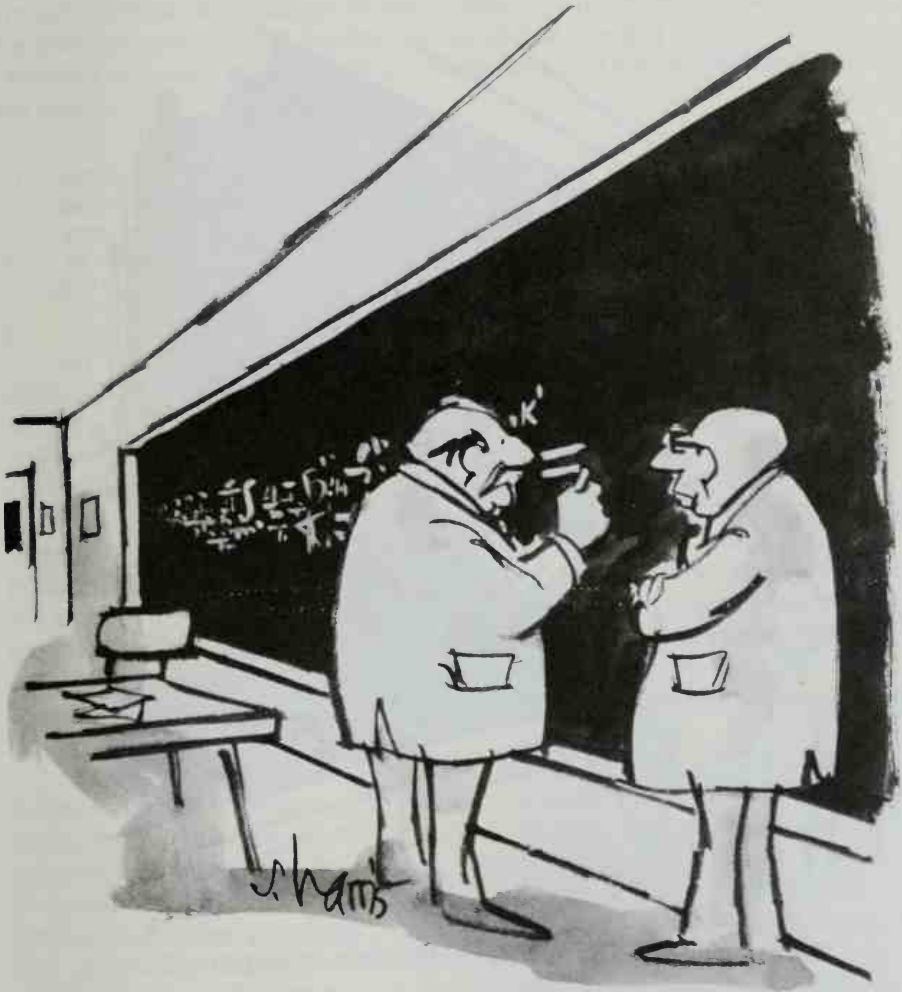
Although it is possible for people to acquire knowledge that makes life worse rather than better (nuclear conflict, made possible by a knowledge of nuclear physics, is an example of a knowledge-based, life-worsening event), Boulding suggests that the usual net result of increased knowledge is improvement in the human condition. Besides, he argues, the growth of knowledge is an evolutionary process, and once it is started there is no turning back. In Boulding's view, then, knowledge may be able to save us, and lack of knowledge may lead us to destruction.

Avenues to Knowledge

Where does knowledge come from? Many of us have succumbed to the mystique of science to the extent that we believe that most knowledge is a result of science. In fact, much knowledge is a result of experience and not the consequence of the rigorous application of the **scientific method**. We all obtain knowledge through daily experience and observation. However, systematic knowledge is generally the product of science and accumulates through careful, well-designed research. A primary purpose for taking a course in social science research methods is to learn how science is done and how scientific knowledge is obtained.

Although not everyone is a scientist, everyone is, to some extent, a researcher. In the words of Kenneth Hoover (1976: 5), "Science is a mode of inquiry that . . . is common to all human beings. Some people specialize in scientific approaches to knowledge, but we are all participants in the scientific way of thinking." There are many other approaches to knowledge, however, such as appeals to authority, tradition, or charismatic leadership. Sometimes people cling to their beliefs in the face of contrary evidence;

ILLUSTRATION 1.1. This cartoon illustrates the scientist's quandary about completing a particular research project or making the findings available.

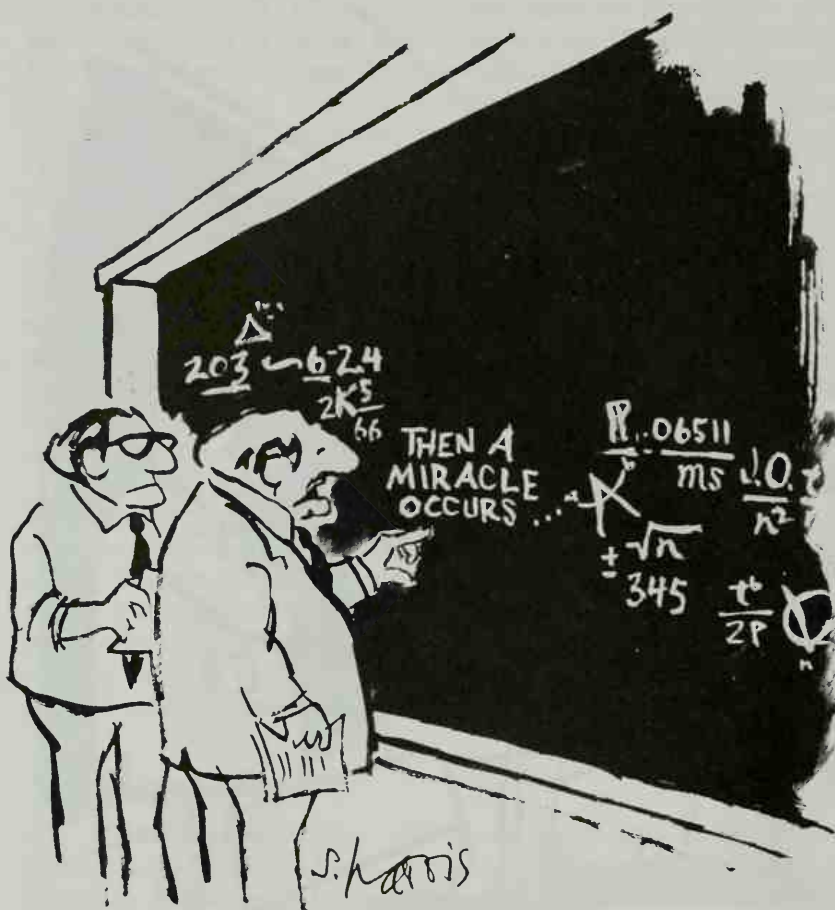


"ON THE OTHER HAND, MY RESPONSIBILITY TO
SOCIETY MAKES ME WANT TO STOP RIGHT HERE."

© Sidney Harris. From *What's So Funny About Science*, Los Altos, Calif.: William Kaufmann, Inc. Reprinted with permission.

perhaps more frequently, they maintain myths about reality when they have evidence neither for nor against the mythic assertions. Research, logical analysis, and systematic comparison are some ways of coping with reality that have been demonstrated to be useful.

ILLUSTRATION 1.2. This cartoon illustrates the importance of researchers' revealing the exact recipe by which they obtained their results.



"I think you should be more explicit here in step two."

© Sidney Harris. From *What's So Funny About Science*, Los Altos, Calif.: William Kaufmann, Inc. Reprinted with permission.

III. SCIENTIFIC METHOD

Part of the value of social research based on a strategy of **verification** and systematic analysis is that it seems to serve a complex society well. The more complex the social environment, the less likely it is that the social cohesion necessary to maintain and sustain the society can be founded on anecdotes, aphorisms, and individual accounts. The highly subjective, and

therefore essentially unshared, individual experience which underlies personal knowledge is replaced by the more disciplined and systematic approach to social knowledge. Hoover (1976: 7) argues that "Knowledge is socially powerful only if it is knowledge that can be put to use," and what seems to convince most people is information that can be double-checked for accuracy.

No one can double-check everything that goes on as the mind deals with the inner feelings, perceptions of experience, and thought processes. Science brings the steps of inquiry out of the mind and into public view so they can be shared as part of the process of accumulating knowledge. (Hoover, 1976: 10)

In other words, the scientific approach is the ultimate democratic approach. It assumes that everyone has a right to the answers. Confronted with the questions, What is so? and How do you know? the scientists are obligated to transmit their knowledge (findings) clearly and often to spell out their implications. In answer to the second question, they are obligated to describe their methods clearly enough that the doubter can follow step-by-step and arrive at his or her own conclusions in the matter. The importance of this essential democratic ethic in science cannot be overstated. Whereas the keepers of the mysteries in other knowledge systems—the priests, the wise ancients—were repositories of sacred, often secret knowledge and rituals, the high priests of science are bound by the ethics of the scientific method to make the "recipes" for their hard-won knowledge public.

The scientific method, at base, is **systematic observation** of nature, followed by telling others of the findings. Scientists are supposed to reason logically, but they do not have a corner on logic. People can begin with assumptions other than those upon which the scientific method is said to be grounded and still deduce, induce, justify, explain, and analyze according to accepted principles of logic. If science includes systematic thought and systematic observation, then part of scientific exploration is really the application of **technology**, of a set of rules or accepted procedures for searching and verifying. What is involved in "scientific" research, then, is careful application of technique, coupled with imagination.

What we must not do, says Ziman (1968), is to distinguish science as a body of knowledge from science as what scientists do, or from science as a social institution. It is only as a unified activity involving all three of these elements, knowledge, technique, and social context, that science can be understood and distinguished from nonscience. The usual definitions of science involve two elements, the scientist and the "Nature" he or she observes. A third element, the scientific community, must also be included for a definition of science to portray accurately what happens. It is signif-

icant that the audience of the scientist's work is not a passive one. The enterprise is corporate both in the doing (in Newton's phrase, one "stands on the shoulders of giants," or builds upon the work of predecessors; also there are typically many contemporaries engaged in similar work standing on those same shoulders) of scientific work and in the public communication process which follows the release of a finding (Ziman, 1968: 8-11).

A widely accepted list of the characteristics of the scientific method (Merton, 1957) includes four basic values: universalism, communism, disinterestedness, and organized skepticism. *Universalism* refers to the idea that scientific truths transcend the person, time, or locations of their discovery; they are presumably relevant or applicable in some more universal context. *Communism* in this context refers to the obligation of scientists to communicate their findings to each other, and to interested people generally, that scientific knowledge is not the property of the individual, school, organization, or nation. *Disinterestedness* refers to the pursuit of scientific knowledge for its own sake rather than for fame, money, power, or other personal profit. *Organized skepticism* is the requirement that scientists not accept other people's findings on faith, but rather check and recheck them before certifying them as accurate statements of what seems to be. One of the applications of this fourth characteristic in the social sciences is a questioning of the view of reality accepted by people generally.

The Assumptions of Science

Sjoberg and Nett (1968: 14) characterize science as "an approach to knowledge that is far more disciplined and calculated than the ordinary inclinations of humans." The scientific method is based upon a number of assumptions which are accepted as "articles of faith" by those who apply it. Among the most important of these "articles of faith" are the following:

1. That there exists a definite order or recurrence of events.
2. That knowledge is superior to ignorance.
3. That a communication tie, based upon sense impressions, exists between the scientist and "external reality" (the so-called "empirical assumption").
4. That there are cause-and-effect relationships within the physical and the social orders.
5. There also are certain "observer" assumptions:
 - a. That the observer is driven to attain knowledge by the desire to ameliorate human conditions.
 - b. That the observer has the capacity to conceptually relate observations and impute meanings to events.
 - c. That society will sustain the observer in his or her pursuit of knowledge. (Sjoberg and Nett, 1968: 23-24)

To conduct research is to express, via one's activity, the faith that seeking answers to questions is better than acting without information.

However, to merit the name "research," one's activities in pursuit of information must proceed according to certain rules. Doing research according to the rules means that one collects and interprets data in line with an accepted set of procedures. The culture of science provides norms which specify how research should be conducted even when the researcher's aims are not strictly "scientific." As presently prescribed, the "proper" procedures ensure that one's assertions about the nature of reality can be checked by other people, that evidence be limited to that accessible to the senses, and that findings be published so that others may check the results. It is the careful application of these standard procedures that transforms unorganized sense impressions into "facts."

Characteristics of Research

The *Random House Unabridged Dictionary* defines research as "diligent and systematic inquiry or investigation into a subject in order to discover or revise facts, theories, applications, etc." In order that it be systematic, rather than random, it necessarily involves a plan. A research plan limits the types of things one will do and the arena in which they will be carried out. When a set of activities are carried out, they have meaning only to the degree that the lessons learned or results achieved are somehow related to prior (and possibly future) activity by asking questions such as "What do these findings mean?" and then communicating the findings intelligibly.

Not all research is "scientific" research. One can be a happy and productive researcher and never contribute to the advancement of scientific knowledge. **Applied research** uses the scientific method to answer a specific question for a specific group at a given point in time and is less concerned with the discovery of new knowledge. For example, a corporation may conduct research to determine why worker morale is low, why employee turnover is high, and why productivity is declining. The research may serve as a case study to illustrate certain principles, but it is not likely to contribute any new scientific knowledge.

The difference between research in the service of science and research in service of other goals lies in the researcher's motivations and objectives. Those who are trying to discover basic laws that seem to govern the social and material world around them are probably "doing science."

Research, whether scientific or not, is a communication process. One of the main reasons for doing research is to get information that will help in decision making. For scientists, the decisions to be made may relate to intellectual propositions which they wish to discard or incorporate into "theory" about how things or people are related. Research of interest to businesspeople may help them make decisions about proposed changes in the company or assess the marketability of a new product. For the educator, research may identify individuals or programs that are not reaching an

expected threshold of achievement. For the organizational analyst, it may locate points of organizational stress or conflict which require special managerial attention.

Most decisions, both personal and organizational, are made without reference to scientific research. Even when research findings are available, they tend to be underutilized. It is one of the assumptions of this book that people would be better off if more of the decisions that affect them were based, at least in part, on relevant research.

If research findings are to be more useful in improving the quality of human life, it is essential for us to be able to distinguish between good and poor research. It is primarily on four dimensions—the degree to which research is systematic, comparative, cumulative, and appropriately communicated—that we need to evaluate the research findings available to us in the course of our occupational roles or encountered in private conversations or in newspapers, magazines, and the electronic media.

IV. LIMITS OF SCIENTIFIC OBSERVATION

We have considered in some detail the nature and limitations of human knowledge because this is a book about how to seek knowledge through research. We have assessed the scientific method as one approach to knowledge and have underscored some of its limitations, because much of the research that is done or that is “consumed” by the public purports to be scientific research. An awareness of the limitations of human knowledge, particularly knowledge grounded in scientific research, is desirable for well-informed people generally and is absolutely critical for people who do social science research or use the results of other people’s research.

The problems of human knowledge we have reviewed are general problems, in that they apply to other modes of acquiring knowledge as well as the scientific method. In much of this book we will urge that some form of systematic observation be used to help people acquire the information they need to make better decisions. However, we do not want to oversell systematic observation as an avenue to knowledge. We can see some of the difficulties attending the search for knowledge of things social by reviewing four basic problems of **observation**: (1) you cannot observe something without changing it; (2) you cannot observe something without misperceiving it; (3) you cannot interpret an observation without misrepresenting it; and (4) you cannot communicate an interpretation of an observation without an additional misrepresentation.

You Cannot Observe Something Without Changing It

For many years social scientists have been aware of **reactive effects**—changes in the behavior of persons being studied due to their awareness of being the objects of study. The fact that observation exerts some influence

on whatever is being observed is sometimes called the **Heisenberg principle**. This principle applies to even inert materials. For example, at the level of atomic physics, the observation of an electron shifts it from its “natural” position and velocity. In social and biological research with human subjects the effects of the Heisenberg principle are even more important in biasing scientific findings. Kenneth Boulding explains:

The famous illustration is that of a man who shouts into the door of a hospital room to his sick friend, “How are you?” The friend says “Fine” and the effort kills him. As we move into the biological and social sciences, the generalized Heisenberg principle—that the attempt to get information out of a system changes it—is increasingly important. Often we cannot investigate living things without killing them. We cannot give a person a questionnaire without changing his opinion. In social systems, a prediction of the future that is believed will change the future itself. (Boulding, 1978: 42)

Reactivity in experimental situations is often called the **Hawthorne effect**, after the classic studies of productivity conducted in the 1930s (Roethlisberger and Dickson, 1939) in which increases in the morale of the persons being studied, increases judged to derive from their awareness that they were part of a special experimental group, were judged to have had more effect upon productivity than any of the variations in work conditions which were the intended factors under study. Reactivity is sometimes called the “guinea pig effect,” and its effects upon research results are often unpredictable.

In 1966 Eugene Webb and his associates published an extraordinarily popular research methods book entitled *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Its main message was that too much social science data were from interviews and questionnaires. A better approach, they argued, was for investigators to use several different methods to attack a single research problem. This approach, which they called **multiple operationism**, was defined as the use of a collection of methods chosen so as not to share the same weakness, and the weakness they were most concerned about was reactivity, or error from the respondent. They identified four different kinds of reactive measurement effect:

1. The guinea pig effect—awareness of being tested changes behavior.
2. Role selection—respondents choose among their many “true” selves according to their definition of the research situation.
3. Measurement as change agent—initial measurement activity introduces real changes in what is being measured.
4. Response sets—people tend to agree rather than disagree, whatever the issue, or they answer in a stereotyped, habitual fashion. (Webb et al., 1966: 13–21)

Researchers try to design projects to minimize these reactive effects, but as long as one uses living respondents who are aware they are being studied, some reactive effect is unavoidable. Webb and his associates ar-

gued that reactivity might be reduced or done away with entirely if researchers drew upon other data sources, such as physical evidence or archival records.

Reactivity is a serious problem in questionnaire and interview research, but it also affects other types of research, such as observational studies of either the participant variety (in which the researcher does what members of the group being studied do, as well as collecting data) or nonparticipant variety (in which the researcher watches and records but does not take part in the activities of the group being studied). Unfortunately, there is evidence that reactivity continues to be something of a problem whatever data-collection procedures are used. Because of the importance of this problem, we will review it several times in the chapters that follow.

You Cannot Observe Something Without Misperceiving It

There are two major types of misperception. One is cultural: we literally do not see many things that our culture, and in particular our language, have not sensitized us to notice. Human perception is selective, and stimuli that catch our attention are ones made salient by our experience. The second source of misperception is a result of the limitations of human faculties of observation. High and low frequency sounds exceed the thresholds of our hearing; certain wavelengths of light are beyond our physiological capacity to perceive. The limitations of our sensory capacities are highlighted by the previously unseen worlds revealed by the inventions created to extend our senses, such as light-gathering telescopes, electron microscopes, radio telescopes, infrared cameras, and X-ray machines. Even the electronic media demonstrate the feeble capacity of the unaided human senses. The air is filled with radio and television signals, but we cannot pick them up unaided; a machine must receive and translate them into the narrow range of visual and auditory signals that we are capable of recognizing.

Technology has vastly increased the dimensions of the world to which we can attend and has raised questions about the dimensions that remain inaccessible. We are not capable of observing the world, the universe, or even human society. Only fragments are accessible to us.

Another reason why observation distorts **external reality** is that it is always incomplete. Although external reality—the universe, if you will—is entirely interconnected, in conducting a research project, we necessarily consider only a portion of it. The act of setting that portion apart, abstracting or isolating it from other variables and influences, creates an artificial situation. Of course the abstraction is essential from the standpoint of the analyst. If he or she must pay attention to everything at once, it will never

be possible to understand anything. Even so, the necessity to truncate reality does not alter the fact that the analyst has changed it.

You Cannot Interpret (Attribute Meaning to) an Observation Without Misrepresenting It

Culture and language facilitate (and impede) both observation and interpretation. First, they may literally affect what is perceived, by sensitizing us to certain stimuli and creating a trained incapacity to pay attention to others. But the abbreviation and distortion are not over when an object or event has been perceived. Interpretation, even more than perception, depends upon the experience and expectations of the perceiver. Kenneth Boulding (1956: 18) has written that the growth of knowledge depends in part upon information received and in part upon an "active internal organizing principle." Such organizing principles are greatly influenced by one's culture. Boulding (1956: 7) makes a distinction between the image (knowledge) and the messages that reach it. His definition of meaning is the change that a message makes in the image. Whether a message, *once received* (remember that because of limitations in our human sensory equipment, many of the messages "sent" are never received) has an impact on the image, and the direction and intensity of that impact, depends upon the personal history and immediate situation of the person receiving the message, as well as on his or her language and culture.

You Cannot Communicate an Interpretation of an Observation Without an Additional Misrepresentation

Communication, whether in writing, speaking, or gesturing, involves a translation from personal to public discourse. A researcher's communication may include a description of what was observed and at least a rudimentary interpretation of what the observation means. Neither is communicated without loss and bias. Communication in any media depends upon the use of symbols with variable meanings and a glossing over of portions of the observation thought to be nonessential or mutually understood. The process of translation into verbal or written language inevitably distorts the observation. There is a corresponding distortion by the receiver of the communication—the one who reads the scientist's log or listens to a verbal presentation—because the reader/hearer does not receive all the messages that are sent and furthermore interprets those that *are* received according to his or her own culture, experience, present situation, and other factors.

To recognize that the human senses, however they may be aided with technological extensions, provide us with data that are incomplete and inaccurate is not grounds for abandoning the search for knowledge. We

may see through a glass darkly, but such sight is probably better than none at all. An awareness of the limitations of human observation and influence may impart an appropriate humility to our research efforts. Even so, there is enough evidence that science “works,” in spite of its limitations, that researchers may proceed with cautious optimism. It is even possible to create techniques that allow us to achieve some control over the variations in individual perception which cloud our sight. An exploration of those techniques is one of the main purposes of this book.

V. RESEARCH AND THEORY

To this point we have said very little about theory. However, as any good student of research methods will quickly learn, theory is virtually always present in the research process. Sometimes we do research to test theory; other times the research **hypotheses** are derived from theory; almost always the **analysis** is greatly strengthened if findings are interpreted in light of available theory. As Robert Merton succinctly observed, it is not enough simply to say that research and theory must be married, “they must not only exchange solemn vows—they must know how to carry on from there. Their reciprocal roles must be clearly defined” (Merton, 1967: 171).

While there are almost as many different definitions of theory as there are writers on the subject, our preference is for one that makes the theory–research link clear and specific. Homans (1964) suggests that theory has three major characteristics: First, it consists of a set of **concepts** or a conceptual scheme. Some of the concepts in a theory are **descriptive** while others are operative. **Operative concepts** are generally referred to as variables, as we will see in Chapter 2. One of the primary purposes of empirical research is to examine relationships that exist between variables that are identified in a theory.

Second, a theory consists of a set of **propositions** that are developed to describe a relationship between the variables. The propositions, if one has developed a **formal theory**, might form a deductive scheme in that some of them can be directly derived from others.

Finally, the theory makes explicit the fact that some of the propositions are, in Homans’ words, contingent in the sense that experience is relevant to their truth or falsity. That is, the propositions are testable in the real world and actual empirical data can be accumulated to determine their **validity**.

What this suggests is that theory can be used to systematize and organize our everyday experiences. From this systematization, we can then derive or develop specific hypotheses that can be submitted to empirical test through the research process. In short, theory can be used to provide order and insight to research activities (Denzin, 1970). Good theories both

organize what we know from prior research and generate propositions that can be tested in future research.

An important aspect of the relationship between theory and research is that research frequently results in the modification of theory. For example, a scientist may have a theory about the relationship between family size (number of children) and the intelligence of the children. The theory argues that the greater the number of children, the less attention the parents are able to direct to each child and this fact limits the child's intellectual development. Data are collected and basically the theory is supported. However, careful analysis reveals that social class is an important factor in the relationship between family size and intelligence. It is noted that children from both large and small well-to-do families have very similar levels of intelligence, as do children from large and small poor families. Further, it is observed that children from the well-to-do families have higher intelligence than do those from the poor ones. The theory must then be modified to take this finding into account, and several alternative hypotheses are generated that can be tested in further research. For example, do well-to-do families provide their children more education experiences, do parents in well-to-do families spend more time with their children, or do children from well-to-do families associate with high achieving peers who pressure them to do well in intellectual activities? A long list of hypotheses linking family size, social class, and intelligence can be generated which can then be tested in research. At times the marriage between theory and research appears to be a never ending circle of theory—research—modification of theory—further research—and so on.

In any research activity the quality of one's work will be improved through a careful consideration of theory. As we present various "how to do it" chapters, the reader must not forget the critical link between the research act—the application of a particular methodology to a research problem—and the body of theory from which that problem may be derived or which can be used to increase understanding and insight during data analysis.

VI. ETHICS OF DOING RESEARCH WITH HUMAN SUBJECTS

In their effort to increase knowledge, social scientists use human beings as their subjects. This fact contributes to the problem of reactivity discussed earlier. A more central problem for the social scientist who studies fellow human beings is that the scientist must constantly be concerned with their welfare. Whereas the laboratory chemist may quite freely expose a research object to any combination of forces in order to observe resultant changes, the social scientist must consider the consequences the research may have on the subjects. Whereas the chemist may wish to see which substance or

combination of substances will change the composition of the compound under study, the social scientist must make sure that his or her research does not result in any permanent change, damage, or injury to the persons studied.

Many of the research examples described in later chapters involve, to some extent, deception of the research subjects, invasion of privacy, and other acts that carry potential harmful consequences. However, we assert that the social scientist must be aware of and respect the dignity, privacy, and worth of other human beings. One cannot justify a course of research that will have serious harmful effects for the subjects of that research, even in the interest of advancing scientific knowledge.

To illustrate the importance of considering such consequences of one's research, we can refer to some examples primarily from medical research. During the 1960s and early 1970s several examples of injury to human subjects in biomedical and behavioral research came to national attention. Perhaps the most infamous, the Tuskegee Syphilis Study conducted by the U.S. Public Health Service, involved withholding treatment from a group of syphilitic black males for 40 years (Brandt, 1978). The purpose of this experiment was to identify a sample of syphilitic men and to observe over time the consequences of untreated syphilis. The researchers did not infect the men with syphilis; that condition was a consequence of the subjects' own behavior. However, once the study was underway, the researchers were active in altering the men's lives and life chances without their awareness or consent.

To minimize the possibility of their subjects' receiving treatment from other medical sources, the researchers gave the syphilitics a free but ineffective treatment (mercurial ointment) with the promise that it would cure their "bad blood" (syphilis). Although effective cures for syphilis were unknown when the study began, medical knowledge about effective treatment increased during the course of the study, and in the late 1940s it was learned that penicillin was a very effective treatment. Nevertheless the scientific objectives of the researchers outweighed their concern for the subjects, and the decision was made to withhold effective treatment from the men. Several procedures were used to ensure that the study was not compromised by the men's receiving penicillin therapy elsewhere. First, the subjects were regularly given free medical exams and told that the ointment was successfully treating their problem. Second, a meeting was held with local black doctors, who were given a list of the 400 men and instructed not to treat them for syphilis but rather to refer them to the Public Health Clinic (the researchers). Third, the Alabama State Health Department was given the same list and requested not to treat the research subjects in their mobile V.D. unit, which regularly came to the county. Finally, several of the men were drafted into the army during World War II and military authorities agreed to cooperate by not treating the syphilitic

men. These procedures to preserve the scientific integrity of the study at the expense of its moral integrity were very successful. When the study was halted in 1972, only 7 percent of the subjects had obtained sufficient penicillin to have had an impact on the disease.

The subjects suffered additional injury because the periodic diagnoses of the disease included a spinal tap. This procedure entailed considerable pain and occasionally caused medical complications. The deception used to motivate the men to submit to such procedures is illustrated in the following portion of a letter sent to them.

Some time ago you were given a thorough examination and since that time we hope you have gotten a great deal of treatment for bad blood. You will now be given your last chance to get a second examination free. This examination is a very special one and after it is finished you will be given a special treatment if it is believed you are in a condition to stand it . . . REMEMBER THIS IS YOUR LAST CHANCE FOR SPECIAL FREE TREATMENT. BE SURE TO MEET THE NURSE. (Brandt, 1978: 24)

The promise of a "special" treatment was sufficient to persuade the men to submit to the spinal tap.

A serious invasion of the subjects' and their families' privacy was also obtained by deception. The research team determined that an autopsy was necessary to chart the effects of long-term syphilis on various organs and tissues of the body. Access to the bodies of the subjects who died was obtained by offering them free burial. Family members did not know that the free burial included an autopsy, and this ploy secured access to the bodies of most of the subjects who died during the project.

Another 200 innocent men suffered injury in the experiment in that although they were free from syphilis they were told that they had the disease. These men were lied to in order to provide a control group suffering comparable psychological stress (knowledge of being syphilitic) for comparisons with the experimental group. They were also given free "treatment" to keep them coming back to the PHS Clinic so that their health could be monitored. When a man in the control group contracted syphilis during the 40 years of the experiment, he was shifted to the experimental group and the effects of the disease observed.

Reports of the experiment first appeared in medical journals in 1936 and continued to appear until 1972. These reports made it clear that the researchers knew the consequences for their subjects of failing to treat the disease. In 1946 it was reported that nearly twice as many syphilitics as controls had died (Brandt, 1978: 25). A similar report in 1955 callously reported that the deaths of over 30 percent of the experimental group were directly attributed to advanced syphilitic lesions (Brandt, 1978: 25).

The study ended when in 1972 an inquisitive journalist exposed it in the national news media. Public pressure forced officials to terminate the

experiment and treat the surviving men. There is no doubt that at least 28 and perhaps over 100 of the 400 subjects died directly from advanced syphilitic lesions (Brandt, 1978: 21). Most of these lives could have been saved and the men's suffering prevented if they had been allowed to obtain effective treatment. In 1972 the Department of Health, Education, and Welfare (DHEW), which includes the Public Health Services, appointed a panel to investigate the ethics of the experiment. The panel concluded that the study was "ethically unjustified" and that the subjects should have been treated with penicillin when it became available (DHEW, 1973).

Another shocking case of the violation of the rights of human subjects was publicly revealed in the early 1960s. Live, human cancer cells were hypodermically injected into 22 elderly patients at Jewish Chronic Disease Hospital in Brooklyn. The patients were deceived about the nature of the experiment, and therefore there was no informed consent. The researchers justified lying to the subjects on the grounds that they "wished to avoid both emotional responses from patients and refusals of participation" (Hershey and Miller, 1976: 7).

Although the most serious examples of injury to research subjects appear in medical research, the behavioral sciences are not without guilt. In an experiment described in detail in the chapter on experimental research, Milgram (1963), a psychologist, led subjects to believe that they had delivered severe electric shocks to another subject. This deception created considerable emotional anguish and guilt and was potentially harmful to the subjects.

Such accounts of a few extreme examples, however, should not suggest that the violation of the rights of experimental subjects is a regular occurrence in medical and behavior research. Rather, these unethical studies stand out because they are unique among the thousands of studies that are conducted each year.

Cardon and his associates (1976) studied the incidence of research-related injuries in biomedical and behavioral research funded by DHEW. A sample of 538 studies was selected from the 2,904 projects funded in 1974 by the National Institute of Health and in 1975 by the Alcohol, Drug Abuse, and Mental Health Administration. The principal investigators of the 538 selected projects were surveyed by telephone interview and an acceptable response rate of 85 percent was obtained. Over half, 331 investigators, said that they had used human subjects. In "therapeutic" research, it turned out that 11 percent of 39,000 subjects had been injured by the experimental treatments. Since many of those subjects were patients with very serious or fatal diseases, the use of more dangerous treatments was said to be justified. Of the 93,000 subjects involved in nontherapeutic research less than 1 percent were reported injured. Also over 95 percent of the injuries in the nontherapeutic projects were trivial, such as soreness from needle insertions. Cardon et al. (1976) concluded that the risk of injury

in a nontherapeutic research project was no greater than the risk a subject encountered in everyday life.

Abuse of human subjects, including the invasion of their privacy, elicited public and governmental concern during the 1960s. In 1966 the Surgeon General issued a policy for the protection of human subjects in biomedical and behavioral research funded by the Department of Health, Education, and Welfare (DHEW, 1971).

In 1971 DHEW published the Institutional Guide on the protection of human subjects. The National Research Act was passed in 1974. This law created the National Commission on Protection of Human Subjects of Biomedical and Behavioral Research and requested institutions sponsoring research to establish Institutional Review Boards (IRBs) to approve projects involving human subjects. IRBs are instructed to review proposed research projects, keeping the following points in mind:

1. To ascertain that the benefits for the subject and the importance of the knowledge to be gained outweigh the risks to the subject. In other words, to determine that the probable benefits justify the risk.
2. To assure that the rights and welfare of subjects are adequately protected.
3. To ensure that legally informed consent is obtained.
4. To provide that the conduct of the research will periodically be reviewed. (Liemohn, 1979)

Informed consent means that the subject or his or her legal representative understands the nature of the study and the risks the subject will be exposed to and then makes a decision to participate free from "force, fraud, deceit, duress, or other forms of constraint or coercion" (Liemohn, 1979: 159). Allowing participation in a social research project to affect a college student's grade is seen by many as exercising undue duress, and this definition of duress has the potential to reduce the availability of college students as research subjects. Providing students with the option of either participating in a research project or writing a term paper brought charges of coercion against a psychology department (Smith, 1977).

Obviously, the use of children in social research requires parental permission in lieu of informed consent. However, the Commission argued that a child seven years of age or older should have the right to refuse participation, because children at that age can usually understand the basic ideas of a research project.

DHEW (1975) identified seven components of adequate informed consent:

1. An accurate description of the research procedures and its purposes must be presented in terms understandable to the potential subject.
2. A description of the risks or of any discomfort or injury reasonably expected must be disclosed.

3. A description of the benefits the subject and the scientific discipline reasonably expect from the research must be noted.
4. Disclosure of any alternative procedures that might be advantageous to the subject must be made. This pertains to biomedical research where patients with specific medical problems are being treated with experimental procedures.
5. An offer to answer questions about any of the research procedures must be given.
6. The instruction that the subject is free to withdraw from the research at any time without prejudice must be fully explained.
7. If there is the possibility of physical injury, an explanation as to whether medical care and compensation are available must be presented.

These procedures are most applicable to experimental research but do apply in varying degrees to other types of research. A controversial recommendation to amend the 1974 act to extend the federal regulation and review requirements to research that is *not* federally funded has been widely discussed in the biomedical and behavioral research communities. Under the proposed policies, universities or private research corporations who receive *any* federal research funds would be required to review *all* research involving human subjects regardless of the source of funds. One noted social scientist commented that "At times it [the proposed regulation] becomes so absurd that it is downright embarrassing" (Pattullo, 1980: 3). He contends that the proposed regulation would require that a political scientist combing the *New York Times* for information about individual politicians would first have to get permission from his university's Institutional Research Board (IRB). Also a sociologist studying the sociology of sports would have to have prior approval before observing and making notes on the level of performance of a major league baseball player in a game.

Federal legislation and regulations governing the use of human subjects are changing. Anyone designing a research project using human subjects should review the policies of the sponsoring organization (university, company, or agency). Almost all will have an IRB familiar with the latest regulations and can assist the researcher to comply with them.

VII. WHY DO RESEARCH?

We conclude this chapter with a brief discussion of why scientists engage in the research enterprise. To answer this question, we must consider the objectives of researchers and how they differ from the objectives of other people. It is sometimes claimed that scientists do research for the satisfaction of getting knowledge; one does research to learn. The personal appreciation of understanding apparently was a major motivation for the

sociologist Robert Park, who described the justification for his embarking on a career as a journalist, and later as a sociologist, in these words:

I conceived a scheme of life that should be devoted to merely seeing and knowing the world without any practical aims whatever. . . . I made up my mind to go in for experience for its own sake, to gather into my soul as Faust somewhere says "all the joys and sorrows of the world." (Raushenbush, 1979: 15)

Another reason for doing research is the hope that the results will help to solve some problem or improve conditions in some way. This justification for research may lead to disillusionment, for many times research findings obtained at great cost are ignored. Nevertheless, the wish that one's research will make the world better continues to provide the psychic and organizational fuel for many studies. Belief in the benefits to be gained from "knowing the facts" was, for example, a primary justification for W.E.B. DuBois's classic (1896-98) survey of black residents of Philadelphia. The explicit purpose of the inquiry was "to furnish local agencies and individuals, interested in improving the condition of the Negro population of Philadelphia, a more comprehensive knowledge of the existing condition of Negroes, so that such work may be directed in the most helpful channels" (DuBois, 1899: xi).

Others assert that continued research—better craftsmanship, more universally applied—is essential to national and organizational survival:

From the earliest civilizations, human beings have required information for survival and social progress. . . . In stark contrast to more primitive times, the survival of today's organization serving human needs depends on a management control system that enables human service managers: (1) To make better *plans*—plans that relate to organizational goals and objectives based on the relative benefits and costs of alternative courses of action. (2) To have better *control*—control that assures efficient and effective action in pursuing the organization's objectives. (Sorensen and Elpers, 1978: 127)

Just as the acquisition of knowledge through research may improve the quality of life for humankind generally, so may the obtaining of knowledge about particular social programs or organizations improve their functioning. Thus we have another reason to do research: to make social organizations work better. The usual name for this kind of research is *evaluation*. Some experts on **evaluative research** echo Boulding's concern that knowledge may sometimes make things worse rather than better. Despite that possibility, Rossi and Williams (1972), like Boulding, are optimistic that increased knowledge will ultimately be a good thing. In their view, social science has had less impact upon contemporary society than it should have, and they claim that "sound evaluative research may allow

social science to make a far greater contribution to society than it has in the past" (Rossi and Williams, 1972: xiv, xvii).

Another reason that people do research is that they find conducting research a satisfying and rewarding activity. Part of the reward may be external to the research process. That is, many scientists "do science" and many researchers "do research" because they are paid to do so. Like plumbers or cabinetmakers, they are craftsmen who live from the plying of their specialized skills and who take some pride in craftsmanship in the exercise of those skills. Conducting social research has become one of the activities of governments and industry—sometimes it is required by law—and when an evaluation or an **impact statement** is needed, one calls in the tradespeople whose training qualify them for the job.

The researcher as craftsman is now a popular image in the research literature. C. Wright Mills is among those responsible for this development. His famous essay, "On Intellectual Craftsmanship," begins with the words, "To the individual social scientist who feels himself a part of the classic tradition, social science is the practice of a craft" (Mills, 1959: 195). Jacob Bronowski (1978: 121) compares the craftsmanship of the scientist to that of the artist and teamster: "The scientist is as completely involved in the whole of his work as any poet or artist and, I suppose, bank manager or truck driver. If he does the job well, it is because it is him." Contemporary textbooks call for "high professional standards" and "competent practitioners" (Hoinville and Jowell, 1978: v) and describe the scientist as "a very special sort of craftsman," who has "a craftsman's intuitive knowledge of these abstract objects" and who possesses "craft skills in the production of data" (Ravetz, 1971: 77, 81). In the chapters that follow, the student-reader will be invited to become a craftsman, to learn and to practice the skills and methods that are used in expanding knowledge about human social behavior.

VIII. SUMMARY

Social science researchers are engaged in a continuing search for "knowledge" about human social behavior. This search is often difficult because the subjects of the research—other human beings—are thinking, feeling, and reacting beings rather than inanimate objects that are little influenced by the decisions and actions of the researcher. While this heightens the challenge of doing good social research, it also increases the satisfaction that comes from research that is well done.

The scientific method of conducting research is based on the principles that scientific knowledge is relevant in some universal context; is freely shared with all interested individuals; is sought for its own sake rather than as a means to fame, money, and power; and finally is checked and

rechecked before being accepted as an accurate statement of what seems to be.

There are limitations to observation of social behavior that must be reckoned with in conducting social research. The Heisenberg principle asserts that observation changes whatever is being observed and this issue is crucial when one is studying human behavior. The researcher's culture and language also influence what is observed, the interpretation of the observations, and the communication of the findings.

The social science researcher operates between two worlds—the everyday world of his or her subjects and the world of concepts and theories that are used to organize, direct, and interpret what is observed in the everyday world. The reciprocal relationship between the world of theory and the world of empirical research is evident in theory guiding research and, in turn, research confirming or modifying theory and suggesting new theoretical propositions which become the focus of further research.

Social researchers must be concerned with the welfare of their subject matter, human beings. Deception, invasion of privacy, exposure to painful experiences, and so on must be used in accordance with accepted canons of research ethics so that permanent physical, social, and emotional injury to the subjects does not occur. Informed consent requires open discussion with potential subjects about the possible benefits of the study as well as possible dangers so that the subjects can make an informed decision to participate or to refuse.

There are many reasons for doing research; the most important is that social research has the potential to improve the quality of human life.

REFERENCES

- BOULDING, KENNETH E., 1956. *The Image: Knowledge in Life and Society*. Ann Arbor: Univ. of Michigan Press.
- 1978. *Ecodynamics: A New Theory of Societal Evolution*. Beverly Hills, Cal.: Sage.
- BRANDT, ALLAN M., 1978. "Racism and Research: The Case of the Tuskegee Syphilis Study." *The Hastings Center Report* (December): 21–29.
- BRONOWSKI, JACOB, 1978. *The Origins of Knowledge and Imagination*. New Haven: Yale University Press.
- CARDON, PHILLIPPE, F. WILLIAM DOMMEL, JR., and ROBERT R. TRUMBLE, 1976. "Injuries to Research Subjects." *The New England Journal of Medicine* 295 (September 16): 650–654.
- DENZIN, NORMAN K., 1970. *The Research Act: A Theoretical Introduction to Sociological Methods*. Chicago: Aldine.
- DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE, 1971. *The Institutional Guide to DHEW Policy on Protection of Human Subjects*. Washington, D.C.: U.S. Government Printing Office.

- . 1973. *Final Report of the Tuskegee Syphilis Study*, Ad Hoc Advisory Panel DHEW, Washington, D.C.: U.S. Government Printing Office.
- . 1975. "Protection of Human Subjects—Technical Amendments." *Federal Register* 40: 11854–11858.
- DUBOIS, W. E. BURGHARDT, 1899. *The Philadelphia Negro*. New York: Benjamin Blom.
- HENLE, ROBERT J., 1969. "Science and the Humanities." Pp. 1–21 in Alden L. Fisher and George B. Murray (eds.), *Philosophy and Science as Modes of Knowing*. New York: Appleton.
- HERSHEY, NATHAN, and ROBERT D. MILLER, 1976. *Human Experimentation and the Law*. Germantown, Md.: Aspen Systems Corporation.
- HOINVILLE, GERALD, and ROGER JOWELL, 1978. *Survey Research Practice*. London: Heinemann Educational Books.
- HOMANS, GEORGE C., 1964. "Contemporary Theory in Sociology." Pp. 951–977 in Robert E. L. Faris (ed.), *Handbook of Modern Sociology*. Chicago: Rand McNally.
- HOOVER, KENNETH R., 1976. *The Elements of Social Scientific Thinking*. New York: St. Martin's Press.
- LIEMOHN, WENDELL, 1979. "Research Involving Human Subjects." *Research Quarterly* 50, No. 2: 157–163.
- MERTON, ROBERT K., 1957. *Essays in Social Theory and Social Structure*. New York: The Free Press.
- . 1967. *On Theoretical Sociology*. New York: The Free Press.
- MILGRAM, STANLEY, 1963. "Behavioral Study of Obedience." *Journal of Abnormal and Social Psychology* 67: 371–378.
- MILLS, C. WRIGHT, 1959. *The Sociological Imagination*. New York: Oxford University Press.
- NEWCOMB, THEODORE M., 1966. "The Interdependence of Social-Psychological Theory and Methods: A Brief Overview." Pp. 1–12 in Leon Festinger and Daniel Katz (eds.), *Research Methods in the Behavioral Sciences*. New York: Holt, Rinehart & Winston.
- PATTULLO, E. L., 1980. "Who Risks What in Social Research?" *IRB* 2 (March): 1–3, 12.
- RAUSHENBUSH, WINIFRED, 1979. *Robert E. Park: Biography of a Sociologist*. Durham, N.C.: Duke University Press.
- RAVETZ, JEROME R., 1971. *Scientific Knowledge and Its Social Problems*. New York: Oxford University Press.
- ROETHLISBERGER, FRITZ J., and W. J. DICKSON, 1939. *Management and the Worker*. Cambridge, Mass.: Harvard University Press.
- ROSSI, PETER H., and WALTER WILLIAMS (eds.), 1972. *Evaluating Social Programs*. New York: Seminar Press.
- SJOBERG, GIDEON, and ROGER NETT, 1968. *A Methodology for Social Research*. New York: Harper & Row.
- SMITH, R. J., 1977. "Electroshock Experiment at Albany Violates Ethics Guidelines." *Science* 198: 383–386.
- SORENSEN, JAMES E., and J. RICHARD ELPERS, 1978. "Developing Information Systems for Human Service Organizations." Pp. 127–172 in C. Clifford Attkisson, William A. Hargreaves, and Mardi J. Horowitz, *Evaluation of Human Service Programs*. New York: Academic Press.
- WEBB, EUGENE, DONALD T. CAMPBELL, RICHARD D. SCHWARTZ, and LEE SECHREST, 1966. *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Chicago: Rand McNally.
- ZIMAN, J. M., 1968. *Public Knowledge*. Cambridge: Cambridge University Press.

Research Design

two

I. Introduction

II. Selecting Problems for Research

Scientific significance as justification for research

Social problems as research topics

III. Types of Research Design

IV. Stages in Research Design: Preparing a Plan

Specifying research objectives

Justifying the project

Utilization of findings

Methods of research

Implementation

V. Principles of Effective Design

Parsimony

Unobtrusiveness

Equal time for analysis and writing

Triangulation

Specialization as trained incapacity

VI. Research Language

Measurement

Variable

Research hypothesis

Operational definition

Reliability

Validity

VII. Summary

I. INTRODUCTION

There are many reasons for doing social research. Sometimes research is motivated by the desire to solve a pressing social problem such as poverty, racial conflict, drug abuse, crime, or dropping out of school. Other times it is a central part of some decision-making process and the research is used to evaluate the potential consequences of alternative courses of action. For example, a company may want to assess the changes in employee job satisfaction that are likely from reorganization of the company. Also sometimes the knowledge gained through research is an end in itself; it becomes a part of our attempt to understand ourselves and the world we live in. Whatever the reason for doing research, its usefulness will ultimately depend upon the quality of the research design.

The design of social research consists of the preparation of a *plan* whereby verifiable knowledge about the research problem is obtained. However, the importance of understanding the process through which scientific knowledge is obtained is not limited just to the scientist but is relevant to almost everyone. Good research training, or at least the ability to distinguish between acceptable and unacceptable research, is a highly valued competency in many professional settings, including city planning and management, business administration, market research, journalism,

the administration of law and justice, social work, educational administration, and health care planning and evaluation.

An understanding of good **research design** is important whether one's interests in research are "**pure**" (knowledge for knowledge's sake) or "**applied**" (knowledge to solve problems). Good research may not always contribute to the body of accumulated knowledge known as "science," but it may help to solve practical problems of communities, corporations, schools, or governments. "Practicality" is often a matter of personal perspective. What is impractical or visionary to one researcher may be pragmatic and utilitarian to another. The important point to remember is that research has little scientific *or* practical value if it is not properly designed.

Our focus in this chapter will be on the design of research in general. Whether data are collected through **interviews**, **questionnaires**, or **participant observation**, or whether research is conducted in a laboratory or in

ILLUSTRATION 2.1. Practicality is often a matter of personal perspective.



"YOU DON'T SEEM TO UNDERSTAND, PRESCOTT. WE'RE NOT TRYING TO CURE DISEASES OCCURRING ONLY IN GUINEA PIGS."

a **field setting**, the principles of good research design apply. We shall spell out these principles, and their application to more specific types of research techniques will be detailed in later chapters.

II. SELECTING PROBLEMS FOR RESEARCH

Before developing a specific research design, one must first decide on the problem to be studied. There is no shortage of researchable problems. Any society consists of far more interactions than can be recorded or analyzed by researchers. The world is full of data; it is the researcher's time and resources that are in short supply. As a result, the critical question to be answered at the outset is not whether the project is interesting or feasible, but whether it is *significant* enough to be worth doing at all. Indeed, the instinct for the significant, as opposed to the trivial, is one of the characteristics that distinguishes the great researcher from the social science hack.

For this reason, people concerned with describing successful strategies of research often cite the early decisions about what to do in a project as most critical. Unsuccessful projects typically suffer from deficiencies in the way research topics are selected or conceptualized. Kenneth Hoover acknowledges the difficulties in identifying a research topic:

The hardest problem in scientific thinking occurs at the very beginning. Once you have solved it, other steps fall into place. This is the problem of limiting the topic, or, more positively, isolating that approach to the topic that will most effectively get at the thing you want to understand. (Hoover, 1976: 42)

It is the posing of apt questions—and aptness is ultimately the utility of the answers the questions call forth—that distinguishes good from poor research, and worthwhile effort from empty **ritualism**.

Scientific Significance as Justification for Research

One way to assess **significance** is to apply the criterion of scientific merit. Unfortunately, there is no sure-fire way to separate the significant questions from the trivial ones. Robert Merton's (1959) important essay "Notes on Problem-Finding in Sociology," published more than two decades ago, remains one of the best treatments of the topic. Merton (1959: xi) says it is hard to find really significant problems: "It is not a matter of dull routine but a difficult task that taxes the imagination." He tries to clarify the necessary procedures by describing what sociological problems are and listing occasions for problem-finding in sociology. The three main parts in the formulation of problems worthy of study, he says, are originating questions (what one wants to know), the rationale (why one wants to know

it), and the specifying questions (queries leading to possible answers in ways that satisfy the rationale).

An important question the researcher should ask is whether the behavior he or she wants to study is what it appears to be. Much effort has been wasted in the explanation of supposed events or conditions that ultimately turned out not to have happened or to have been described inaccurately. It is thus essential that the researcher know *what* is to be studied before attempting to explain *why* it occurs or exists. Simple description—the characterization of what is—is a crucial foundation of scientific explanation. Another essential type of originating question is to ask about relationships between classes of **variables**, either within particular institutions or in a variety of institutional settings.

Merton (1959: xix) pointed out that “the bare question does not constitute the problem.” It is not enough to simply ask “Why?” There must also be an explanation of why the answer to the question is worth having. Possible legitimate rationales are idle curiosity (knowledge for its own sake) and practical curiosity (knowledge for the sake of making life better). Sometimes a double rationale is possible, when the answer to a question both advances the accumulation of systematic knowledge and also has practical benefits. A study having such double relevance is the work of Condie (1976), which specifically explored ways to get people to donate blood (practical use) but also shed light on theories of voluntary action and altruism (benefit to systematic knowledge). Often the most significant theoretical questions have relevance to many problems and institutions.

The reason for emphasizing the scientific rationale, says Merton, is that questions can be too easily multiplied. The requirement that one must demonstrate that a question deserves attention prevents scientists from being swamped by too many questions: “The requirement of a rationale curbs the flow of scientifically trivial questions and enlarges the share of consequential ones” (Merton, 1959: xix).

Not only is the world full of data, it is also full of potential questions. Researchers can ask sufficient Whos? Whats? and Wheres? in a few minutes to account for a lifetime of scientific effort. The point is that good research is more likely to happen when the right question has been asked.

To summarize, a scientific problem is not merely “a question put to nature” but is a question that incorporates a plan for achieving an answer (Ravetz, 1971: 132–133). A scientific problem involves a sequence of actions rather than a piece of finished knowledge disconnected from the process by which it was obtained. “Unless there is some idea of how the work will be done,” says Ravetz (1971: 133), “there is no way of knowing whether the solution can even be achieved; and in general the form that the tentative solution takes will depend on the projected means of its accomplishment.” In other words, posing a scientific problem necessarily includes notions,

half-formed though they may be, of the kinds of research design that may yield the desired answers.

Social Problems as Research Topics

At least three kinds of problems stimulate social research: policy problems, problems of social philosophy, and problems intrinsic to developing scientific disciplines (Greer, 1979: 49). Issues related to the last two are usually cast in the language of abstract intellectual discourse. The first, however, is the arena of the common person. It is here that we deal with what is wrong with the community, the society, or the world.

Policy problems are the practical, urgent problems a society defines as its own. They reflect the values of society. However, the urgency of a problem according to some abstract, external standard is not necessarily related to the urgency accorded it by society. Social problems appear and disappear by popular definition. According to a recent textbook on social problems, "no condition, no matter how dramatic or shocking to someone else, is a social problem *unless and until* the values of a considerable number of people within the society define it as a problem" (Horton and Leslie, 1981: 5).

When researchers draw their research questions from the popular scheme of societal problems, they run the risk that they will pay more attention than is merited to trivial or spurious problems or overlook crucial variables that do not happen to be part of the contemporary framework of a problem (Manis, 1974: 305-306). Moreover, the organization of research around simplistic, catchy slogans is apt to deflect resources from efforts with higher probabilities of success.

For example, in the late 1960s the "population explosion" was defined by many as one of the critical social problems facing the United States and the world. Today, a decade later, the U.S. birthrate is the lowest in history and the population crisis has disappeared. During that period when population growth was defined as a dominant national problem, a few like demographer Ben Wattenberg cautioned that the "Explosionists'" explanations of society's ills directed attention away from the real problems.

But what is wrong, and dangerous, and foolhardy is to make population a crisis. Doing so will simply allow too many politicians to take their eyes off the ball. When Explosionists say, as they do, that crime, riots, and urban problems are caused by "the population explosion," it is just too easy for politicians to agree and say sure, let's stop having so many babies, instead of saying let's get to work on the real urban problems of this nation. (Wattenberg, 1972: 28-29)

Two of the present authors argued that too hasty closure about what was really "the problem" might ultimately result in ill-conceived and wasteful research:

In the end, the energies, idealism, and resources devoted to altering people's ideas about family size in order to produce zero population growth, particularly in the more developed societies, may be in vain. It is not that they will not succeed; but even if zero population growth is achieved, many of the problems which supposedly derive from population size will still exist. Life is short; human energy and talent are limited; it is tragic that solutions to many of our most pressing problems will be postponed because despite the neo-Malthusian ideologies, the variance in social conditions attributable to population size alone proves to be so small. If our assessment of the evidence is correct, millions of well-meaning, talented people will have been caught up in a movement which focused on the wrong variable. (Bahr, Chadwick, and Thomas, 1972: 10-11)

The point of all this is that the researcher had best be cautious about accepting the prevailing definition of a social problem as the framework within which research is to be conducted. An awareness of the pitfalls and biases inherent in accepted definitions of social problems may help the researcher achieve some degree of objectivity.

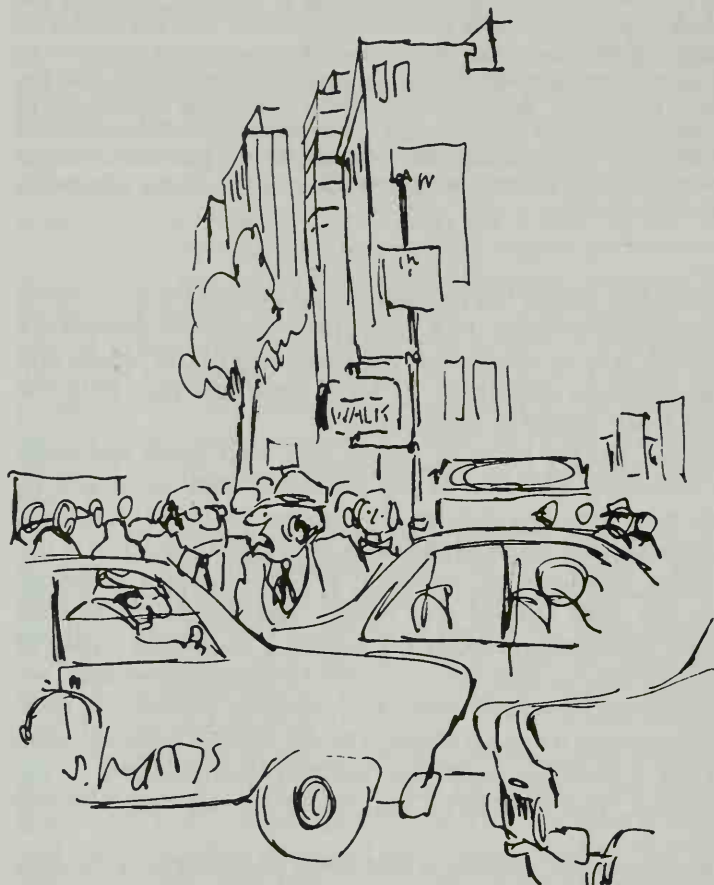
There is another danger faced by researchers who study the social problems of the moment, and that is the likelihood that more will be expected of them than they can deliver. As Caplow (1971: 12) has said, one of the great fallacies that masquerades as a sociological principle is the notion that all social problems are solvable. If people expect social researchers to come up with "things that work," and the problems they confront are chronic rather than readily resolvable, a continued outlay of resources which does not produce the anticipated (though unrealistic) solutions may lead to disenchantment with social research. Indeed, the massive cuts in federal spending for social research by the Reagan administration may represent such a backlash. Since the Kennedy administration, the social researchers have "had their day," and things are not demonstrably better.

Caplow's recommended solution is that social researchers avoid such boomerang effects by designing projects that have a chance of succeeding because they follow correct design principles. His advice about how to design effective projects of social improvement also applies to projects of social research:

. . . the difficulty lies . . . in our unwillingness to apply existing sociological knowledge to the real world in, precisely, a serious and rational way. . . . We must try to learn how to distinguish between feasible projects of social improvement (which may not succeed) and defective projects which *cannot* succeed because they lack essential parts. . . .

It is very likely that we already have enough sociological knowledge to improve American society beyond recognition if we can only learn to design feasible projects, as distinct from the platforms and slogans to which we now pin our hopes of social improvement. (Caplow, 1973: 32-33)

ILLUSTRATION 2.2. The researcher had best be cautious about accepting the prevailing definition of a social problem as the framework within which research is to be conducted.



“Today’s problems should have been solved in the 1950s, but in the 50s we were solving the problems of the 20s, in the 20s we were solving the problems of the 1890s . . .”

© Sidney Harris. From *What’s So Funny About Science*, Los Altos, Calif.: William Kaufmann, Inc. Reprinted by permission.

III. TYPES OF RESEARCH DESIGN

The continua along which research projects may be classified are named by reference to the labels attached to the poles of the continua. First, there is the *quantitative–qualitative* dimension, which refers to the degree of precision of measurement used in a project. It is important to remember that

neither **quantitative** nor **qualitative** measurement is "good" in itself. We agree with Filstead (1970: vii) that researchers should use methods appropriate to the topic at hand and that complex measuring devices may become ends in themselves and therefore impediments to knowledge rather than intermediate tools which enhance understanding. Everyone understands that some measurement instruments are too precise, too highly calibrated, for some tasks. One does not measure the distance between cities in centimeters, and for most household measurements a ruler or a yardstick, not a micrometer, is sufficient.

The dimension underlying the most common classification of types of social research is the *method of data collection*. Although data collection is only one stage of research design, people often label an entire project on the basis of the way the data are gathered. Thus, a project in which the source of data is interviews is called an **interview study**, and one in which the researcher lives among his or her informants and shares their lifestyle is called a **participant observation study**. A project in which the researcher manipulates one factor and watches the effect on another is labeled an **experiment**.

Research projects may also be classified in regard to their *primary objectives*. If a project is aimed at evaluating the effectiveness of some organization or program, then it is **evaluation research**, whatever the sources of the data used in the evaluation. Other types of design based on project objectives include the **hypothesis-testing** study, **comparative** research (Holt and Turner, 1970), **descriptive** research, and environmental or **social impact** research. In our view, inasmuch as data collection is only one of many stages in a research design, and high quality studies frequently use data obtained in two or more ways, it is more useful to classify studies by the project objective than by the method of data collection.

Another continuum along which research designs are categorized is the *time-orientation* of a project. If data collection is focused at a single point in time, a study is **cross-sectional** in nature; if the same respondents or subjects are observed at several points in time, then the design is said to be **longitudinal**. As with many of the design continua, there is a composite or "compromise" type of design, the **retrospective**, in which respondents questioned at a single point in time attempt to reconstruct their actions at various earlier times.

Still another way of identifying types of research design is to consider *whether the data were collected to specifically test the research question* or whether data collected for some other reason are used. If one is collecting or creating data specifically for the purposes of the present research, the project is referred to as **primary data collection**. On the other hand, when an enterprising researcher finds an existing data set which can be applied to his or her own purposes, the project is referred to as **secondary analysis**. Of

course a secondary analysis may be conducted on data originally gathered by interview, questionnaire, simple observation, participant observation, experiment, census enumeration, or whatever.

Finally, research designs may be classified in terms of the *degree to which they impinge on the lives of informants or subjects*. Thus, research that is conducted in ways that bother people little or not at all is called **unobtrusive** (nonreactive), in contrast to the more typical experiment, interview, or questionnaire designs which require the cooperation of subjects. It should be apparent that most observational studies and secondary analyses are unobtrusive.

IV. STAGES IN RESEARCH DESIGN: PREPARING A PLAN

Having selected or been assigned a problem, and having determined the type of design one wants to use, the researcher faces a multitude of additional questions. How should the research problem be conceptualized? Has there been any previous work on the topic, and if so, what findings have resulted from that work? How might data relevant to the problem be collected, and why is one mode of data collection preferred over others? How are the data to be analyzed, and in what form will the findings be released?

The best way to deal with these questions is to write a **research proposal**, which is actually a design committed to paper. A proposal is a statement of one's research ideas and plans, complete with justification and explanations about how decisions among alternative design possibilities were made. Good research is systematic, comparative, cumulative, and communicated, and the preparation of a written proposal forces the designer to face the hard questions, make explicit the projected comparisons, show how the work will be cumulative, and lay out in stages precisely how it will be conducted.

A research project may be large or small, may involve a single investigator or a team of researchers, may consist of a series of complicated tests of an entire theoretical framework or a simple test of hypotheses. Like any project, from the building of a birdhouse to the orchestration of a political campaign, there are certain things that need to be checked along the way. That is, there is a set of rules, based on experience in previous projects, that governs the management of successful projects, whatever their scale or content. A research project has several parts, and the nature of the research partially determines what those parts are and how they are arranged. However, there are certain parts that are essential to *any* project if it is to operate efficiently.

A *scientific* research project, says Carlo Lastrucci (1967: 55), can be broken down into eight stages:

1. The formulation of the problem in a way that empirically testable propositions or hypotheses are produced.
2. The study of related literature for relevant data or methods.
3. Construction of a research design and its rationale.
4. Determination of the universe (context, locale, population) and of the size and characteristics of the sample of units to be studied.
5. Data collection and processing into workable form.
6. The interpretation of the data, that is, what are the findings?
7. Verification of the interpretation through confirming or questioning of other studies, or by confirming or rejecting the original hypothesis.
8. The presentation of findings in a report.

Theodore Caplow (1971: 40) reduces the stages of scientific research into five, which he says represent the simplest form of a project:

1. Planning the project (Lastrucci's stages 1 and 2).
2. Designing procedures (Lastrucci's stages 3 and 4).
3. Gathering data.
4. Analyzing data (Lastrucci's stages 6 and 7).
5. Reporting results.

Sometimes one or more of the five essential stages is ignored, short-circuited, or stunted. Even in the most carefully planned project, things go wrong. The odds that something serious will go wrong which may reduce the project's capacity to produce usable and defensible findings increases with every omission or abbreviation of the essential stages.

Let us now present a more detailed outline of topics that ought to be included in a comprehensive research plan. It is recognized that no general outline will suit all needs and that more extensive documentation and justification is required for some projects than for others. Nevertheless, use of the following guidelines will help to ensure that essential steps of procedure are not overlooked. Included in these guidelines are issues relevant to both pure and applied research.

Specifying Research Objectives

The first step in developing a research plan will usually be to describe the nature of the information that the project is intended to produce. That is, the researcher should specify the basic descriptive, theoretical, or administrative questions that are to be addressed and indicate what information is required to answer them. General research objectives might include answering questions like the following: What is the nature of the problem that is to be addressed? Is there a relationship between two or more variables? What actions might be taken that will alleviate the problem or lead

to the attainment of goals? What results have occurred as a function of the implementation of some program?

In addition to identifying general research objectives, the researcher will usually want to specify specific research objectives which the project will address in achieving its general objective. Specific research objectives might describe how answers to the research questions that are posed will provide the information needed to address the theoretical, practical, or administrative questions that prompted the research question.

Justifying the Project

Justification typically involves explaining why the project is needed in terms of both the priority of the problem and the extent of previous work on it. For example, the researcher might want to address such important questions as the following: What research has already been done on this topic? How does the proposed project relate to that research? Is adequate information already available from previous research or other sources? Is it essential that a new study be done or can existing information be used? How important is the problem, policy, or program addressed by this project? What are the criteria of importance applied in making this judgment? What specific decisions or actions could be based on the findings of this project? Who might make these decisions? What is the potential impact (benefits and/or costs over the foreseeable future) of possible decisions based on the findings of this project? Who, specifically, will experience these impacts? What are the probable costs of proceeding *without* the information that could be provided by this project? Who, specifically, will bear these costs?

Utilization of Findings

This step will usually pertain to how findings of the project may be used and whether anyone is responsible for developing specific plans for utilization. Important questions that the researcher might want to address are the following: Who will interpret the data and findings from the project and formulate conclusions and recommendations? How and to whom will the findings and recommendations be presented? What kind of reports (interim, final) are anticipated, and on what schedule? What audiences will be addressed? How and to whom will a report be made describing specific actions taken and indicating whether recommendations were implemented? Will the data and findings be available for secondary analysis?

Methods of Research

In this step, the procedural steps that will be used to accomplish the goals and objectives of the project will be explained in detail. This explanation should include a justification for the methods selected and should

demonstrate concern for an appropriate matching of objectives, methods, and available resources. Standard research procedures may need to be adapted to the specific purposes of the project. The researcher will want to provide more specific information on the following types of questions: What are the units to be studied? What characteristics of the units are of interest? What general type of design (case study, participant observation study, interview study, experiment, etc.) is proposed? If the effects of change in one or more characteristics are being investigated, how will possible alternative explanations for the observed changes be dealt with?

The researcher will also be interested in specifying **sampling** methods that are to be used, what specific **instruments** (tests, scales, indexes) will be used, how those who collect the data will be trained, and so on. Other important questions include what procedures will be used in preparing the data for analysis and what specific analytical techniques will be used.

Implementation

This final step involves specifying how the actual research procedures will be implemented. Who is in charge of the project? What other persons will participate? What specific roles or tasks will each perform? How will time and personnel be allotted to each stage of the work? This step will also deal with the specification of the project budget and the allocation of that budget to specific research tasks.

V. PRINCIPLES OF EFFECTIVE DESIGN

We have described the major components of a research design and the order in which they are usually arranged. Now let us enumerate some general principles—standards, if you like—which may be applied to improve the design of research.

Parsimony

The first principle is akin to the scientific law of **parsimony**, or to Occam's famous razor. It is more often urged upon philosophers and theorists than upon researchers, but the consequences of ignoring it may be even more serious in the world of real systems where the researchers must operate than in the abstract world of the mathematician or philosopher. William of Occam was a fourteenth-century English philosopher who urged that the causes advanced as explanations for phenomena be as few and as simple as possible. The application of Occam's razor in scientific explanation produces "**elegance**," or discovery involving the fewest possible factors. If the razor is carefully applied, its user emerges with a "simple" hypothesis or explanation.

Doing research usually involves creating a system. Even if the researcher is his or her own assistant, typist, clerk, and data processor, it is essential that the activities associated with these essential roles be done in an orderly fashion, with the product of each of the stages of the research process related appropriately to each of the other parts. Moreover, the system created by the researcher is generally an applied system, functioning in the world of people and things, rather than an abstract mental system.

Research problems can usually be minimized by strenuous efforts to keep the systems we create for collecting, analyzing, and interpreting our data as simple as possible. In practice, this means keeping the numbers down: no more investigators, assistants, clerks, coders, or computer programmers than absolutely necessary. Generally, it means reducing fieldwork to a minimum and eliminating many of the "wouldn't it be interesting to know" questions in the interest of creating a brief, effective research instrument. It means using multiple methods to collect the data necessary to validate and authenticate our results but not collecting data for which there is little probability of use. Finally, it means being able to justify each of the operations imbedded in a research design, and being willing to cut unmercifully those that cannot be demonstrated as essential. To restate the first basic principle of effective research design: *Keep the system simple.*

Unobtrusiveness

A second principle, related to the first, has to do with limiting the expensive and often problematic forays into the field to occasions when they are truly necessary. Primary data collection is very costly. Fieldwork is obtrusive: it bothers people and may create ill will. Furthermore, obstruction produces reactive effects. That is, the interaction with people created in the course of our data collection may create circumstances that compromise the data being collected. Such threats to validity are known as "reactive measurement effects." Among the "reactive measurement effects" are the four kinds of error from the respondent that may contaminate one's findings, including the guinea pig effect, the role selection effect, the effect of measurement as a change agent, and the operation of response sets or patterns of "yea-saying" or "nay-saying" (Webb et al., 1966: 13-21).

There are also reactive effects which may be described as errors from the investigators, all of which will receive more extensive treatment in later chapters. These include interviewer effects (the characteristics of an interviewer affect the data he or she collects in many ways, most of them outside the control of the investigator) and the effects of changes in the research instrument (the interviewer or research instrument may not be the same at one time as it was in a previous time). Human observers have fluctuating adaptation levels and response thresholds; they increase or decrease in

competence and fatigue, and an experiment may change in subtle ways as the experimenters become more experienced or bored (Webb et al., 1966: 21–23).

The upshot of all of this is that there are many sources of error in fieldwork, whether experimental, observational, or survey. The competent social researcher will need to take all of these into account. Some can be resolved in part by devising methods of data collection that minimize the need to create artificial situations. Much of the success of Webb et al.'s *Unobtrusive Measures* stems from the appeal of their creative approaches to nonreactive data collection. One famous sociologist of religion, in conversation with the authors, referred to primary data collection as "trips to the rockpile" and said that he avoided such trips whenever possible. In a subsequent chapter we shall discuss the many types of **unobtrusive data** and how they can be successfully exploited. Here, let us merely state the design principle: *Where possible, avoid obtrusive data collection or, if it is unavoidable, be aware of its pitfalls.*

Equal Time for Analysis and Writing

A third principle of effective research design has to do with the allocation of resources to projects. We single it out for special mention because so many research projects suffer from the violation of this principle. Most projects use too large a share of project resources in data collection and then have insufficient time or money for an adequate, comprehensive analysis, interpretation, and write-up of the findings. There are some constraints in obtaining research funds which heighten the problem. External funding is more readily obtainable for projects that collect large masses of data than for projects in which thinking and writing are the dominant activities. Researchers whose livelihood depends upon a succession of grants may find their time consumed by data collection and by the preparation of proposals for projects to collect more data, and there may be little financial payoff in the reprocessing of previously collected, underanalyzed data.

The fact that we sometimes describe an entire research design in terms of the data collection effort may be another indicator that the essential work of the project—the core of social research—is often seen to be the collection of data rather than the analysis and interpretation of the results. Caplow's (1971: 40) prescription to prevent the shortchanging of essential design stages is that the time allotted to a project be divided evenly among the five stages of planning, designing procedures, gathering data, analyzing data, and reporting results. However, not all projects will require that uniform allocation of resources. Our own experience suggests a simpler rule of thumb: allow as long for analysis, interpretation, and write-up as for all other stages combined. The principle may be briefly stated this way:

Thinking and writing about data are as important as data collection and deserve at least equal time.

Triangulation

A fourth principle of good research design may be summarized in three words: *triangulate when possible*. Generally, **triangulation** refers to the search for consistency of findings from different observers, observing instruments, methods of observation, times, places, and research situations. Triangulation embraces the methods of replication and includes the practices usually followed to estimate the **validity** and **reliability** of research findings.

Research findings are affected by the nature of the research method applied, by the researcher's specific knowledge of that method and how it should be applied, by the personal and professional characteristics of the researcher, including mood, idiosyncracies, and perspectives, and finally by the unique and ever-changing characteristics of the phenomena being observed (Denzin, 1970: 298–299). Triangulation in all its possible varieties permits the researcher to escape some of the variance attributable to these four sources of discrepant observations.

One of the most comprehensive definitions of triangulation is Norman Denzin's (1970: 301–310) treatment of four types of triangulation: *data*, *investigator*, *theoretical*, and *methodological*. **Data triangulation** refers to the focusing of data from different settings upon a single problem. According to Denzin, there are three main kinds of data triangulation: time, space, and person. One who time-triangulates will use data sets collected at different times; one who space-triangulates will use data from a variety of locations; and one who person-triangulates will use data from two or more levels of aggregation, namely from individuals, groups such as a family or work group, and collectivities or communities.

Investigator triangulation refers to the use of multiple rather than individual observers. The concept can be applied at every stage of the research process: two interviewers may record different accounts of the same interview, and two analysts confronting the same computer output may interpret it differently. A finding that many investigators agree about is more convincing than findings vouched for by a solitary investigator.

Theory triangulation is Denzin's (1970: 303–306) term for the assessment of a single observation or data set from the standpoint of several theoretical perspectives. In essence, one investigator approaches a data set with multiple conceptual perspectives, or several investigators, each with a distinctive perspective, may view and interpret the same data set.

What Denzin refers to as **methodological triangulation** applies to the use of variations in the measurement process, or to the use of different measurement processes. The former is often called **within-method trian-**

gulation; an example is the inclusion of two different self-esteem scales (two different measuring instruments) in a single questionnaire. A **between-method triangulation** refers to measures of a single characteristic or relationship obtained in two or more different modes of data collection. For example, one would have more confidence in a researcher's conclusion that a population was in fact characterized by low self-esteem if measures of self-esteem drawn from questionnaires, from content analysis of biographical statements, and from the reports of trained observers all revealed markedly low levels of self-esteem in comparison to similar measurements obtained from appropriate comparison groups. This "between method" approach to triangulation is the one emphasized by Webb et al. (1966) in their classic *Unobtrusive Measures*.

This principle may be simply summarized as: *Triangulate when possible.*

Specialization as Trained Incapacity

Our final general principle of research design is related to triangulation, but it deserves special mention because it can be so inherent in the various social research specialties that researchers may be unaware of it. This is the "law of the hammer," which is the propensity of a child encountering a hammer to suddenly find that everything in the world needs pounding. The professional training of researchers concentrates their attentions, and ultimately their skills, on one or a few research techniques. The universities turn out survey researchers who prescribe survey research as the response to any informational problem, experimentalists who eschew all "knowledge" not grounded in experimental design, ethnomethodologists who decry quantitative research as artificial, incomprehensible, and error-prone, and small-group specialists who invariably recast organizational questions in forms that can be dealt with by panels of staff from different administrative levels.

These specialists in different methodological approaches also tend to be grounded in different theoretical perspectives. The net result is that while we tend to give lip service to **multimethod approaches**, most research continues to be monomethod. People who make decisions about research design often select a mode of data collection before they have worked through the assumptions and implications of a research design sufficiently to make an informed decision regarding methods.

Martin Trow has captured the essence of this principle in a memorable paragraph:

Every cobbler thinks leather is the only thing. Most social scientists . . . have their favorite methods with which they are familiar and have some skill in using. And I suspect we mostly choose to investigate problems that seem vulnerable to attack through these methods. But we should at least try to be less parochial than cobblers. Let us be done with the arguments of "participant

observation" *versus* interviewing—as we have largely dispensed with the arguments for psychology *versus* sociology—and get on with the business of attacking our problems with the widest array of conceptual and methodological tools that we possess and they demand. This does not preclude discussion and debate regarding the relative usefulness of different methods for the study of specific problems or types of problems. But that is very different from the assertion of the general and inherent superiority of one method over another on the basis of some intrinsic qualities it presumably possesses. (Trow, 1957: 35)

The principle is that there are no methods that are intrinsically better than others; the utility of a method is always situation- and problem-specific. Researchers who specialize in one or two research strategies are especially liable to too-early reification of design and likely to err in the direction of their own research skill. An operative rule might be that researchers consider data collection in their own specialties only after the data-collection strategies in which they are least prepared have been considered and ruled inappropriate. But perhaps we ask too much in suggesting that a specialist function in a manner that minimizes, even temporarily, the influence of his or her specialty. Let us restate the principle in positive terms and hope for the best: *Heed the law of the hammer.*

VI. RESEARCH LANGUAGE

We will conclude this chapter with a brief overview of the language of research, highlighting the typical concepts that researchers use to describe their activities.

Measurement

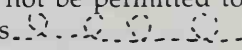
The central problem of most research methods is **measurement**. Measurement implies comparisons of some kind, which are the stuff that give meaning to everyday reality. For example, we might want to know if the crime rate in our community has gone up or down, if members of different social classes hold unique attitudes toward social welfare programs, if religious attitudes affect the probability of a person's engaging in socially disapproved behavior. All of these questions imply that there is something out there that the researcher is going to "measure."

However, measurement in a research context is not always clear and straightforward. When someone says, "It is hot today," they apparently mean that, in comparison with some unspecified other day, today's temperature is high. But the statement is ambiguous. The day or period against which "today" is being compared is not identified; it might refer to the same date last year, last week, yesterday, or to some abstract average. Moreover, we have no way of knowing that one person's definition of "hot" is in any

way calibrated with another's. The speaker's definition of "hot" may begin at about 80 degrees Fahrenheit; the hearer's may start at 95 degrees.

Much of the ambiguity in the statement is a result of the low level of measurement involved. The adjective "hot" is generally presumed to be one end of a **continuum** consisting of several points, such as "cold," "cool," "warm," and "hot." This four-point continuum may be extended to 6, or 10, or 15 categories by the appropriate use of modifiers such as "very," "quite," "moderately," and "slightly." The everyday conversationalist never explains how many points there are on his or her temperature continuum, and hence the statement, "It's hot today," is essentially uninterpretable. We don't know whether "hot" is the category next to the highest bearable temperature, or somewhere between "pleasant" and "burning."

Now let us see how communication is improved by the use of a measurement scale, say the Fahrenheit scale of temperature. To one familiar with the scale, the statement "It is 32 degrees" communicates a discrete, interpretable message independent of the observer's personal tolerance for heat or cold. In other words, even though the scale of measurement is simply an agreed-upon convention, it increases the likelihood that communication about the weather, people's activities, or the general situation of mankind will be intelligible to others. Kenneth Hoover (1976: 28) has put it nicely: "Measurement, properly conceived and executed, at least has the potential for reducing the ambiguity of words and sentences. At the same time, improperly conceived measurement is dangerous precisely because it can be so powerful." It is powerful because people believe and policy makers act upon the presumably more accurate information a reported measurement conveys.

The use of a standard of measurement—say, a yardstick for measuring length—also implies some norms about how that standard will be applied. For example, if one is measuring the length of the Eastern seaboard, the prevailing norms of measurement presumably would require that the yardstick be laid end-to-end in a straight line thusly . The person doing the measurement would normally not be permitted to angle the yardstick in such a way that recurring loops  produced a much longer "length of seaboard" than would be derived from straight-line measurement. However, the standards of measurement—for example, whether angling off into loops is permitted—are themselves socially determined. We agree upon a standard as well as upon a set of more or less implicit rules about how that standard is to be applied.

Variable

In research, the thing that we are interested in measuring is usually referred to as a variable. For example, crime rates and attitudes toward social welfare programs can be viewed as variables.

Many research projects begin as a desire to describe the intensity or frequency of some behavior. For example, a later chapter will discuss a research project that was designed to describe the academic achievement of minority students attending public schools in the United States (Coleman et al., 1966). The behavior being studied, in this case, the academic achievement of minority students, can be referred to as a **dependent variable** because it is assumed that it is "caused by" or associated with some other variable or set of variables that the researcher is interested in examining.

Independent variables are those factors that explain or predict the dependent variable. Differences in facilities, curriculum, and the training and preparation of faculty and administrative personnel might be suggested as independent variables which have an effect on the dependent variable of academic achievement. Minority group relations theory suggests that segregation into all-minority schools might be another independent variable that has an important influence on academic achievement.

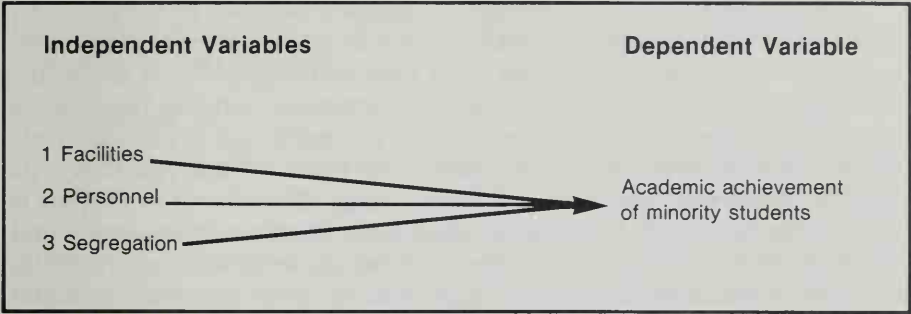
The term *variable* implies that the behavior or phenomenon studied varies from zero or some low level to some higher level. Academic achievement, the dependent variable, could vary, for example, from a very low ACT score of 5 to a very high ACT score of 30 or better. Quality of school facilities could vary from an old dilapidated school with a small crowded library and no science laboratory to a new modern school with a large library and a well-equipped science laboratory. Personnel could vary from inexperienced teachers with only bachelor's degrees, to highly experienced teachers with master's degrees and even a few with doctorates. The important thing to remember is that variables are what one is interested in measuring in research. They are things or events in our social world that can take on different values.

Research Hypothesis

In its simplest sense, a research hypothesis suggests a particular relationship between a dependent variable and one or more independent variables. For example, one might hypothesize that low-quality educational facilities and poorly trained teachers might contribute to lower levels of academic achievement among minority students. Possible hypothesized relationships between several independent variables and a dependent variable are outlined in Illustration 2.3.

Research hypotheses are what the researcher is interested in testing. He or she collects data on the dependent and independent variables and tests the relationship between the two. Hypotheses may be derived from a variety of sources but they typically come from one's theory or are suggested by experiences or observations of a given social behavior.

ILLUSTRATION 2.3. Relationship between public school facilities, personnel, and segregation and academic achievement of minority students.



Operational Definition

An **operational definition** is a specific set of instructions explaining how a variable is measured. The operational definition must be sufficiently clear and specific so that readers can understand how the variable was measured and whether it was a good indicator of what was being studied. An operational definition must be sufficiently specific that other researchers can, if they desire, replicate the research.

Connotative definitions found in dictionaries are not operational definitions. For example, a dictionary definition of racial discrimination is “a showing of partiality or prejudice in treatment; specific., action or policies directed against the welfare of minority groups” (*Webster’s New World Dictionary*, 1970: 403). This is not an operational definition as it does not explain how to measure racial discrimination. If observational techniques were to be used by the researcher to determine how much discrimination occurred among students in an integrated high school, the operational definition would specify exactly which behaviors the observer would count as discrimination. The list could include, among other things, the refusal to sit by a minority student in class or the lunch room, voting to refuse membership in social clubs to minority students, damaging minority students’ books, calling minority students names, and physically assaulting minority students. The bias of measuring discrimination by observation is that such behavior is often committed in secrecy. If the researcher used hidden cameras, one-way mirrors, and similar covert observational strategies, reasonably good estimates about the extent of discrimination in the high school could be obtained.

Operational definitions may also be developed for use in the analysis of records of agencies, organizations, and companies. In our example, racial discrimination could be defined as the number of interracial fights reported to school officials. Vital statistics, as well as information in arrest records,

court records, school records, and similar records are frequently used as operational definitions of research variables.

Many researchers use self-report to define a variable. In our example, high school students could be asked, either in an interview or a questionnaire, how often in the past two years they had engaged in a list of discriminatory acts. The list should be comprehensive, ranging from subtle avoidance to physical assaults. In addition, it is important to include a time frame, such as how often in the past "two years," "one year," or "six months" have you: (list of discriminatory acts). The bias of self-report is that respondents may be afraid to admit their discriminatory activity and the researcher ends up with a gross underestimate of such behavior. Or at times certain individuals may take pride in their discrimination and inflate their accomplishments to impress the researcher.

The researcher must decide whether to use observation, records, self-report, or a combination of these strategies to measure a variable. As was stressed earlier in this chapter, triangulation or the use of two or more techniques helps compensate for the biases of a given measurement technique.

An example of the difference in the research results created by different operational definitions is a study of "Hidden Delinquency" (Murphy, Shirley, and Witmer, 1946). The research team for this project studied the delinquent behavior of 114 boys, age 11 to 16 years, in a lower-class neighborhood for a five-year period. They first operationally defined juvenile delinquency as the number of court complaints issued against the 114 boys during the five years. Records revealed that the 114 boys had committed 95 acts that had received official court attention during the five-year period. The same 114 boys were also observed by case workers during the five years and were informally interviewed by them. The case workers were asked to give conservative estimates of the delinquent actions of the boys. The case workers reported that the 114 subjects had committed 6,416 delinquent acts. The estimation of delinquency obtained by these two different operational definitions varied a great deal, 95 versus 6,416. It is suspected that "reality" was somewhere between the two estimates. The courts had no doubt missed many delinquent acts, and perhaps the case workers had exaggerated the boys' delinquency. This example illustrates the importance of operational definitions to the results obtained in research.

Reliability

An important characteristic of the measurement of a variable or a research finding is its reliability. Reliability refers to consistency or stability of the measurement of a variable using a given operational definition. In the above example, if two or more observers independently observed the same classroom or lunch room using the same operational definition of

discrimination, they should have reasonably similar counts of this behavior. An observer using the same operational definition in two or more identical or nearly identical classrooms should obtain very similar measures of discrimination. A number of specific strategies to increase reliability will be presented in later chapters.

Validity

Validity, another important characteristic of an operational definition, is the degree to which an operational definition actually measured what it was supposed to measure. In other words, is the measurement a "real" indicator of the theoretical variable in question? In the discrimination example, one concern about validity might be whether the behaviors observed in the high school were examples of racial discrimination or simply personal conflicts unrelated to race.

Validity is extremely difficult to assess. The few studies that have attempted to measure validity have done so by comparing different indicators of the same theoretical variable. Quite often the different operational definitions have produced rather different estimates of the variable. For example, during World War II a sample of people who, according to government records, had redeemed war bonds were interviewed and asked whether they had done so (Hyman, 1944). Seventeen percent of those whose records indicated they had cashed in bonds, an act that was generally defined as unpatriotic, denied that they had done so. The large difference between the self-report and the supposedly correct records casts doubts on the validity of the self-reported bond cashing behavior. Validity is a serious concern to social scientists, especially when they are studying socially disapproved actions.

Several specific techniques to assess and enhance validity will be discussed in subsequent chapters. Generally such techniques either tend to be extremely difficult to apply or are inconclusive in demonstrating validity. A widely used technique is to use **multiple indicators** as was done in the war bond and delinquency studies. Unfortunately, multiple indicators are not usually available for many variables social scientists study. Even when they are, it may not be possible to determine which of the different indicators is the most valid or the closest to reality. In some cases, the difference between multiple operational definitions is very small and the researcher feels confident that either measure is a valid indicator of a particular variable. Whatever technique is used to assess validity, the final judgment rests with the researcher. In effect, he or she must satisfy himself or herself that the operational definitions are valid and then proceed.

Other concepts that are central to the language of research include problems of **sampling** and assessing **causality**. Because of the importance

of sampling, it will receive separate treatment in the next chapter. The question of assessing causality will be treated throughout the text as specific research methods are discussed.

VII. SUMMARY

We do research for many reasons. Among the most important of these are to attempt to find solutions to important social problems, to improve our decision-making ability, and to increase our knowledge about ourselves and our world. Whatever the motivation, however, the ultimate utility of one's research will depend upon the quality of the research design.

To talk about research design is to talk about a plan. This plan includes a clear specification of the question that one wants to answer, the procedures one will use in collecting data relevant to that question, and the approach to be used in analyzing and interpreting that data so as to shed light on the question.

We typically assess the significance of our research problems by determining their scientific merit (such as by asking why the answer to the question is worth knowing at all), or by determining their utility in solving practical, urgent problems of our society. It must be cautioned, however, that the connection that exists between urgency of a problem as defined by society or some of its members and the urgency of that problem as determined by some objective, external standard is often weak. Because of this the researcher must be careful about being diverted toward trivial or spurious problems simply because they attract public attention.

There are a variety of different kinds of research projects. For example, projects might be classified according to their primary objective—they may focus on evaluation, hypothesis testing, description, comparison, or impact assessment. Projects also vary along a time continuum: cross-sectional studies focus on data collected at a single time, longitudinal studies with data from several points in time, and retrospective studies attempt to reconstruct the past to better understand present conditions. Studies that collect data specifically for purposes of the present research are referred to as primary; if the research uses an already existing data set, the project is referred to as secondary analysis. Studies also vary in terms of the degree to which they impinge on the lives of the research subjects or on their degree of obtrusiveness.

A successful research project will generally include a series of steps or stages. These include formulating a researchable problem, constructing a research design and its rationale, gathering the data, analyzing and interpreting the data, and reporting one's results. A well-developed research plan will also give attention to (1) clearly specifying the research objectives, (2) providing adequate justification for doing the research, (3) specifying

how the research findings will be utilized, (4) discussing the specific methods of data collection, and (5) describing how the actual research procedures will be implemented.

Five key principles or standards of research design have been identified. These are parsimony, unobtrusiveness, allowing time for analysis and writing, triangulation, and selecting the right tool for the problem at hand. A researcher who follows these five principles will avoid many of the pitfalls of doing research.

As is true for any person who practices a trade or skill, the researcher needs to understand the specialized language of research. Concepts that will be used again and again in the chapters that follow include measurement, dependent and independent variables, research hypotheses, operational definitions, reliability, and validity. A clear understanding of these concepts is an essential step in understanding research design.

REFERENCES

- BAHR, HOWARD M., BRUCE A. CHADWICK, and DARWIN L. THOMAS, 1972. *Population, Resources, and the Future: Non-Malthusian Perspectives*. Provo, Utah: Brigham Young University Press.
- CAPLOW, THEODORE, 1971. *Elementary Sociology*. Englewood Cliffs, N.J.: Prentice-Hall.
- 1973. "Toward Social Hope: II—The Workable Plan." *Columbia Forum* 2 (Spring): 28–33.
- COLEMAN, JAMES S., et al. 1966. *Equality of Educational Opportunity*. Washington, D.C.: U.S. Government Printing Office.
- CONDIE, SPENCER J., W. KEITH WARNER, and DAVID C. GILLMAN, 1976. "Getting Blood from Collective Turnips: Volunteer Donation in Mass Blood Drives." *Journal of Applied Psychology* 61, No. 3: 290–294.
- DENZIN, NORMAN K., 1970. *The Research Act*. Chicago: Aldine.
- FILSTEAD, WILLIAM J. (ed.), 1970. *Qualitative Methodology: Firsthand Involvement with the Social World*. Chicago: Markham.
- GREER, SCOTT, 1979. "On the Selection of Problems." Pp. 48–52 in John Bynner and Keith M. Stribley, *Social Research: Principles and Procedures*. New York: Longham.
- HOLT, ROBERT T., and JOHN E. TURNER, 1970. *The Methodology of Comparative Research*. New York: Free Press (Free Press Paperback Edition, 1972).
- HOOVER, KENNETH R., 1976. *The Elements of Social Scientific Thinking*. New York: St. Martin's Press.
- HORTON, PAUL B., and GERALD R. LESLIE, 1981. *The Sociology of Social Problems*, 7th ed. Englewood Cliffs, N.J.: Prentice-Hall.
- HYMAN, HERBERT, 1944. "Do They Tell the Truth?" *Public Opinion Quarterly* 8: 557–559.
- LASTRUCCI, CARLO L., 1967. *The Scientific Approach: Basic Principles of the Scientific Method*. Cambridge, Mass.: Schenkman.

- MANIS, JEROME G., 1974. "The Concept of Social Problems: Vox Populi and Sociological Analysis." *Social Problems* 21, No. 3: 305-315.
- MERTON, ROBERT K., 1959. "Notes on Problem-Finding in Sociology." Pp. ix-xxxiv in Robert K. Merton, Leonard Broom, and Leonard S. Cottrell, Jr. (eds.), *Sociology Today: Problems and Prospects*. New York: Basic Books.
- MURPHY, FRED J., MARY M. SHIRLEY, and HELEN L. WITMER, 1946. "Incidence of Hidden Delinquency." *American Journal of Orthopsychiatry* 16 (October): 686-696.
- RAVETZ, JEROME R., 1971. *Scientific Knowledge and Its Social Problems*. New York: Oxford University Press.
- TROW, MARTIN, 1957. "Comment on Participant Observation and Interviewing: A Comparison." *Human Organization* 16: 33-35.
- WATTENBERG, BEN, 1972. "The Nonsense Explosion: Overpopulation as a Crisis Issue." Pp. 19-32 in Howard M. Bahr, Bruce A. Chadwick, and Darwin L. Thomas (eds.), *Population, Resources, and the Future: Non-Malthusian Perspectives*. Provo, Utah: Brigham Young University Press.
- WEBB, EUGENE J., et al. 1966. *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Chicago: Rand McNally.
- WEBSTER'S NEW WORLD DICTIONARY, 1970. *Webster's New World Dictionary*, 2nd College Ed. Englewood Cliffs, N.J.: Prentice-Hall.

Sampling

three

I. Introduction

II. Random Sampling

Unrestricted sample

Restricted sample

Systematic sample

Stratified sample

*Disproportionate stratified
sample*

Cluster sample

Area sample

III. Nonrandom sampling

Convenience sample

Purposive sample

Snowball sample

Quota sample

IV. Sample Size

V. Summary

I. INTRODUCTION

Often the goals of social research projects are to describe selected characteristics of a population or to look for patterns in the relationship between characteristics. When the population being studied is large, it is often impossible, and certainly inefficient, to study every person in that population. A study of marital satisfaction in American society that attempted to observe, interview, or send a questionnaire to every married person would cost more than the results could possibly warrant. And imagine trying to involve every married couple in the United States in a laboratory experiment! Fortunately it is not necessary to question every person in a large population to arrive at accurate estimates of the results that would be obtained if it were possible to interview everyone. The use of an appropriate **sample** of married people allows us to describe the level of marital satisfaction in the United States or to test the relationship between stress and marital satisfaction almost as accurately as if we had studied every married person. Thus, scientific sampling makes it possible for the researcher to describe a population or to test a hypothesis on a relatively few research subjects and yet generalize the findings to the larger population.

However, if the sample is to represent a larger population accurately, then that sample must be drawn very carefully, according to a strict set of rules. The rules of sampling have been carefully developed by statisticians

and researchers so that relatively small samples reflect the characteristics of large populations quite accurately. For example, for many types of problems a properly drawn sample of 2,000 is sufficient to represent the entire population of the United States.

Sampling is most frequently used in survey studies, but the same logic applies to selecting experimental subjects, documents for content analysis, or people or places to observe. The units sampled may be individuals, families, organizations, books, television programs, census tracts, towns, classrooms, schools, counties, or states.

To describe procedures of sampling, let us first introduce a few definitions. The term **universe** usually refers to the total group or population the researcher wishes to study. For example, all married Americans would constitute the universe for the study of marital satisfaction in American society. **Population** is sometimes used synonymously with universe; at other times, it refers to a more clearly identified portion or sample of the universe of possible respondents or units.

The **sampling unit** is the individual element or person that, when combined with other elements or persons, makes up the total population to be studied (Mendenhall, Ott, and Sheaffer, 1971). The sampling units may be groups of individuals, families, clubs, school classes, census areas, organizations, companies, baseball teams, city governments, books, media programs, cars on a freeway, minutes in an hour, and so on. A list of the sampling units is called a **sampling frame** and it is from this list that the sample is drawn. Ideally a sampling frame would contain an entry for every unit in the population of units to be studied, but usually it represents only a portion of the universe. The more accurately it represents the total universe, the better the sampling frame is. It is important that the sampling frame be as complete as possible.

A **parameter** is a characteristic of a population. For example, if 75 percent of all married people in the United States are in fact "very satisfied" with their marriage, then this is a parameter describing the population of married people. A *statistic* is a characteristic of a sample that is used to estimate a parameter of a population. In the example above, a statistic describing a sample of married people may be that 73 percent are "very satisfied" with their marriage. The statistic (73 percent) obtained from the sample is a reasonable estimate of the population parameter (75 percent).

II. RANDOM SAMPLING

The usual strategy to obtain a representative sample is to draw a probability sample. The process of doing this is called **random sampling**; it refers to the selection of units from a universe or population so that every unit has exactly the same chance or probability of being included in the sample

(Kish, 1965). For example, if a researcher wished to select a random sample of 100 families from a village with 650 families, he or she might first make a list of the names or addresses of the 650 families. Each name or address would be assigned a number. Then 100 numbers between 1 and 650 would be chosen by chance, so that each unit had a 1 to 6.5 (or 100 to 650) probability of being selected. One way to do this would be to take the first 100 numbers between 1 and 650 from a table of random numbers and the correspondingly numbered units on the list would constitute the sample. An alternative way to select the sample would be to number 650 ping-pong balls, mix them in a bowl, and then pick 100 balls from the bowl. Whatever method was used, the 100 selected random numbers or ping-pong balls would then be matched with the names on the list to provide a sample. Each of the 650 units on the sampling frame would have had the identical probability of being chosen, and therefore this random sample should represent the entire village of 650 families. It is essential that the selection process be random, and that the probability of being selected be precisely the same for all units. If selection is done on some other basis such as selecting families whose homes are readily accessible on paved roads, avoiding houses with vicious dogs in the yard, or neglecting homes with locked chain-link fences around them, then the total population is not accurately represented. Homes with large dogs have a zero probability of inclusion, and those closer to the center of town may be several times more likely to be picked than the homes hidden in the trees in the outlying areas.

The first step in drawing a random sample is to clearly identify the population and decide on the sampling unit. When these have been identified, then the sampling frame—a “master list” of units—must somehow be obtained or approximated. In this age of computers and social security numbers it would seem that every conceivable population would have a list of members, but that is not so. For example, small towns rarely have a roster of residents; neither do many big cities, counties, or states. There are some attempts to prepare such rosters: city directories are published annually by commercial firms, but such directories are usually incomplete and out-of-date. On the other hand, public schools, junior colleges, and universities have rosters of students, and many religious, civic, and professional organizations have membership rolls.

Despite their inherent imperfections—not everyone is listed, especially people who move a lot—city directories are excellent sampling frames. The most popular city directory is prepared by R. L. Polk and Company of Detroit for over 1,400 cities and suburban areas. A Polk Directory lists all of the known adult residents of the city by name and address. In addition, Polk Directories include valuable information about the people listed, including occupation, employer, and spouse’s name if the person is married. This information is obtained by a door-to-door canvass which is updated annually, and the canvassing is supplemented by employee rosters

obtained from major employers. In a study of social change in Middletown (Muncie, Indiana) the city directory allowed us to select samples of specific populations including one sample of women and another of intact (husband-wife) families. Unfortunately, the number of cities for which directories are prepared is limited.

Telephone directories are frequently used as a sampling frame because of their ready availability. Also, most telephone books are updated annually. Telephone directories can be aggregated to provide a sampling frame for a county, a group of counties, a state, or even several states. The largest collection of telephone directories compiled for a sampling frame by the authors was for a study of divorce in eight intermountain states: Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming (Albrecht, Bahr, and Goodman, 1983). The telephone directories covering these eight states were purchased from the various regional telephone offices and were combined to provide a sampling frame for the eight-state region.

Two major problems with using telephone directories as sampling frames are that not everyone has a telephone, and that some people have unlisted numbers. Poor people and the more transient tend not to have telephones and thus are underrepresented in the telephone book. Groves and Kahn (1979) estimated that in 1975 between 8 and 10 percent of U.S. households did not have a telephone. They confirmed that low-income minority populations, rural people, poor people generally, and renters had fewer telephones per household than other people. However, differences are small, and therefore bias introduced by using telephone directories as sampling frames is manageable unless one is interested primarily in low-income or disadvantaged ethnic populations.

Unlisted numbers are a more serious problem. It is estimated that as many as one-fifth of all residential telephones were not listed in current directories in 1974 (Glasser and Metzgar, 1975). Moreover, unlisted telephone numbers are more frequent in metropolitan areas; estimates for Detroit are 35 percent and for San Francisco, 40 percent (Groves and Kahn, 1979: 20). On the other hand, Sudman (1973) estimated the percentage of unlisted telephones in Chicago at 21 percent, which is near the national average. Apparently the reason some people do not list their phone numbers is to reduce nuisance calls, such as telephone soliciting or obscene calls. However, most of these numbers are unlisted because individuals or families have recently moved and were assigned numbers after the annual directory was published. Thus, the longer it has been since the latest telephone directory was published, the more the underrepresentation of new arrivals. Therefore, unlisted numbers can introduce a serious bias, especially in metropolitan areas and especially in the last few months before a new directory is published. Nevertheless, Perry (1968: 695) argues that usually telephone directories are adequate sampling frames.

Records of utility companies and tax rolls are occasionally used as sampling frames. Almost every household has electrical power and is on the county or state tax rolls. The bias with using either of these sources of sampling frames is that renters—again, typically lower-income and more transient segments of the population—are frequently omitted. The landlord or property owner is listed on the tax rolls and often, especially for apartments and other multifamily complexes, the electrical power and other utility bills are not in the name of the temporary occupant. Thus people who rent an apartment, condominium, or duplex might be absent from the sampling frame. This bias is more severe for tax and utility company records than for telephone directories.

When a sampling frame is not available, the researcher may have to create one. One technique is to combine several incomplete lists of the study population. Duplicates (names appearing on two or more lists) are eliminated, and one ends up with a more complete sampling frame than that available from any single source. For example, in the early 1970s two of the authors conducted a study of the adjustment of American Indians in Seattle, Washington (Bahr, Chadwick, and Stauss, 1972; Chadwick and Stauss, 1975). According to the 1970 census there were about 8,000 American Indians dispersed throughout Seattle, but no list of Indian residents was available. Accordingly, we decided to combine the membership and client rolls of several Indian organizations. Client rolls of community social agencies, arrest records, and public school enrollment records were searched for the names and addresses of American Indians. The subscription list of a weekly Indian newspaper was also obtained. The organizations that cooperated by sharing membership or client lists did so under the conditions that the master list would be confidential and used only for research. The names from the various lists were compiled into a single master list and duplicates eliminated. The resulting sampling frame included 3,000 American Indian adults. It was estimated that spouses and children of these 3,000 adults brought the total number of Indians represented in the list close to 8,000, the total population reported by the census.

Another way to construct a sampling frame for special populations is to use a screening device to locate members of the group. For example, we were interested in studying the process of divorce among ever-divorced residents in eight western states (Albrecht and Kunz, 1980). Telephone directories served as an initial sampling frame and a very brief questionnaire was mailed to a random sample of over 1,000 households. The "screening" instrument identified potential respondents who had at any time been divorced and also provided some essential background data on the general population. The major data-collection process targeted those potential respondents who identified themselves as "ever-divorced" in the screening instrument.

It should be clear by now that sampling frames containing every

member of the population are rare. Instead, the researcher is usually forced to settle for a more-or-less complete frame. If the researcher is fortunate, he or she has some idea about the nature and extent of bias in an incomplete sampling frame. City directories and telephone directories are often adequate sampling frames. Utility company records and property tax rolls may be good lists of dwelling units but are less representative of households generally. Often it is necessary to build one's own sampling frame by compiling several incomplete lists or by screening a larger population for members of the group to be studied.

Given a sampling frame, there are many ways to draw random samples. We will consider seven of the most popular modes of sample selection.

Unrestricted Sample

The defining characteristic of an **unrestricted sample** is that when a unit has been selected for the sample, it is returned to the frame. This means that a given unit (individual, family, household, etc.) may be included more than once in the sample. The reason for replacing units this way is that the probability of any given unit being drawn remains identical for the entire sample. For example, if we were to draw a sample of 25 students from a class of 100 students, the first one selected would have a $1/100$ chance of being selected. If the first student is not replaced, then in the second selection each unit has a $1/99$ chance, in the third $1/98$ chance, and the student selected last has a $1/76$ probability of being included in the sample. The difference between $1/100$ and $1/76$ is considerable and violates the premise of random sampling.

An unrestricted sample of students in a junior high school will serve to illustrate this mode of sampling. Suppose the three grades have a total of 300 students and a sample of 50 students is to be selected. The 300 names on the class rolls would be numbered 1 to 300 and 50 numbers within this range would be randomly selected from a table of random numbers (see Illustration 3.1). This table of random numbers would allow for a given number to be repeated, and therefore each student will have a $50/300$ or $1/6$ chance of being assigned to the sample.

Restricted Sample

In a **restricted sample** the selected units are not replaced. A restricted random sample is generally used when the universe is large enough that nonreplacement changes the odds of a single unit's being drawn only very slightly. For example, if a sample of 1,000 is selected from a population of 500,000, the difference in the probability of selection between the first unit ($1/500,000$) and the last unit drawn ($1/499,000$) is so small that it can be ignored. The choice between an unrestricted and restricted sample is an important one only when the population is relatively small.

ILLUSTRATION 3.1 How to use a table of random numbers.

12752	18264	34818	84528	53807	52091	48276	40836	30528	35647
69124	43923	41416	73187	16380	19590	10122	14418	35407	32684
97435	60454	65807	70753	52253	76740	30158	70286	10289	99163
92373	61770	29261	19627	74414	39841	59315	37325	90112	64741
67441	41970	44852	21380	34607	55222	59867	82199	84395	76574
99883	40135	11859	19937	52888	87894	91372	67183	70745	39819
92210	54886	79425	22573	90612	20518	67595	80907	97086	74350
62318	14759	17690	13303	50608	43920	28048	83005	15597	16539
88858	64298	76064	97701	31627	70453	88069	24336	23393	51331
47448	42702	49180	40756	21917	74699	20938	33330	51539	49262
51721	86963	96292	15079	73843	77350	19506	90889	49774	60646
45904	39610	44516	40608	63003	52540	58215	96426	94622	99893
57759	27516	45258	63906	16110	11509	54061	80777	38117	81705
97173	67688	21568	10216	87179	21139	12220	13066	98413	62882
71575	83508	76873	19664	96125	79772	13502	33060	26846	83531
29568	95603	77664	95315	92915	99650	71665	23133	73813	94684
33783	20537	39175	91327	48727	60403	83877	89630	92888	60657
17943	51746	98986	78202	68337	16205	62192	87307	94114	88920
86495	28690	63679	73860	90050	95561	72909	67407	58254	52862

A section of Appendix A, a Table of Random Numbers, is presented above to use in this example. If we want to draw a sample of 50 students out of a junior high school with 300 students, the following procedures would be followed.

1. The size of the population is 300, so the numbers selected must be three digits. One- and two-digit numbers such as 1, 17, or 83 are preceded by the appropriate number of zeros: 001,017,083.

2. The table has columns of 5-digit numbers and we must decide which three digits to use. We can arbitrarily decide to use the right-hand three, the middle three, or the left-hand three. In the first column in the table above, the three right-hand digits would be 752, the middle three 275, and the left-hand three 127. We must decide which three to use and then must be consistent as the sample is selected. In this example we will use the middle three digits.

3. The number can be selected by starting at any point: the top, bottom, left-hand side, right-hand side, etc. Neither does the direction we move through the table matter as long as we are consistent. In this example we will start at the lower left-hand corner and proceed up the column. The first number is 649, the second 794, the third 378, the fourth 956, and the fifth 157. We will continue through the table until 50 numbers have been selected between 1 and 300. When we hit the top of a column, we will move to the bottom of the next one and proceed up it.

4. Three-digit numbers greater than 300 will appear, as in the example we had to go to the fifth number, 157, to find one less than 301. The numbers 301 through 999 will be ignored.

5. It is possible that the same number will appear more than once. The second or third time a number appears, it is ignored and the researcher continues until 50 unique numbers between 1 and 300 have been selected.

An example would be the selection of 500 students from a large university. The campus' directory, listing the names of 20,000 students, might serve as the sampling frame. Typically, the names in the directory would be numbered and a table of random numbers would then be used to select 500 numbers between 1 and 20,000. Any numbers repeated in the table

would be ignored, so that effectively replacement does not occur. Thus, the first student has a .00005 ($1/20,000$) chance of being selected and the last student a .000051 ($1/19,500$) chance. The difference is trivial, and thus the researcher need not bother with replacement procedures.

Systematic Sample

A **systematic sample** is a convenient way to draw a sample from a large population when a printed list is available to use as a sampling frame. In systematic sampling every n th name is selected from the list (see Illustration 3.2). City directories and telephone books are alphabetized sampling frames of households or families in a community that can be used for systematic random samples. The first step is to count the number of units in the sampling frame. When the directory is large, the researcher can sample a few pages and count the average number of pages to determine the number of names in the directory. The estimated number of names in the frame is then divided by the number of units to be included in the sample to determine the interval between names. For example, if the sampling frame was 20,000 names and a sample of 1,000 is desired, then every 20th name would be selected.

It is important that the systematic sample be randomly started. In the example above, a number between 1 and 20 would be randomly selected (probably by using a table of random numbers). If the random starting number was 14, the researcher would select the 14th name on the sampling frame, and then every 20th name thereafter. The sampling would not stop when 1,000 names had been chosen but would continue to the end of the list, even though the sample increased beyond 1,000. The advantage of a systematic random sample is that it is quicker and easier to draw than an unrestricted sample.

An example of systematic sampling was the use of the city directory to select a sample of men in Muncie, Indiana, as part of the Middletown study. The 1976 city directory contained 584 pages of individual listings. The average number of men's names to appear on a page was 83.2 and thus we calculated that the directory contained the names of 48,589 men. Since we desired a sample of 500, we divided the population by this number to determine that we needed to select every 97th name. Pieces of paper were numbered 1 to 97, mixed in a box, and one (37) was drawn out. The 37th name in the directory was selected and then every 97th name thereafter. A sample of 521 men was obtained using these procedures.

Stratified Sample

A **stratified sample** is used when the researcher wants to ensure that a certain segment of the population is represented in the sample. The population is divided into subgroups called strata, and independent sam-

ILLUSTRATION 3.2. A systematic random sample.

A E S Systems 381 W 2230 N	373-0498	Aarons Super Service Appliance	
A G CABINETS & FURNITURE		Repair Co 269 W 400 N	375-0389
255 S Mountain Way Dr Orem	224-6250	Aase Mary Ann 362 E 700 N	374-8468
A I D S 854 S State Orem	226-8891	Abacus Computer Services 240 E Center	377-0076
AIM Distributing 379 N University Av	375-2644	Abadan Bookkeeping Service	
A J Michaels & Co 379 N University Av	374-6895	110 S 500 W	375-6184
A & K Cleaners 1250 N State	375-2611	Abad Santos Aurora 931 E 100 N	375-7725
ALP Systems 190 W 800 N	375-0090	Abarca Rolando 380 S 700 W	377-9436
ALPS-AUTOMATED LANGUAGE PROCESSING SYSTEMS		Abbasi Reva 1715 S 500 E Orem	226-0695
190 W 800 N	375-0090	Abbey Cappel Of Orem 1497 S State Orem	225-2489
A M Real Estate Inc 90 N 100 E	373-8855	Abbey Carpet Of Provo 455 S 900 W	375-6128
AMCO Automatic Transmissions		Abbey Medical	
810 W Columbia Ln	374-8847	44 E 600 S Salt Lake City - Orem Tel No-	224-9808
A M F ELECTRIC CO INC		ABBEY WEDDING COUNSELORS	
1030 N 520 E Orem	224-4330	91 N 700 W	373-0846
AMFAC Drug Supply Co		Abbot David & Sherry Nibley Hall	374-5091
1364 S 300 W Salt Lake City Ut		Abbott Andrea 669 E 800 N	377-2506
Toll Free-Dial 1 & Then	800 662-4217	Abbott Charles F 3737 Foothill Dr	224-6508
AMOCO 794 N State Orem	224-6333	Patricia Nelson Cynthia Lisa Alyson	
A P C Data Centers 1525 W 820 N	373-8518	3737 Foothill Dr	225-6396
ARIX ENGINEERS ARCHITECTS & PLANNERS 1815 S State Orem	225-8491	ABBOTT CHARLES F atty	
ART FURNITURE MFG		2696 N University Av	375-8900
936 N Geneva Rd Orem	225-4658	Abbott David 1466 S 280 E Orem	224-1369
ARTU Inc 1918 N 90 W	224-2594	Abbott David C 675 N 850 W	377-0649
A S C O EQUIPMENT CO		Abbott John D 525 N 650 E Orem	225-6588
1845 W 500 S Salt Lake City	973-8828	Abbott Kristine O-Whitney Hall	377-9376
A T R A TRANSMISSIONS		Abbott Lw 642 N 500 E	374-8738
855 S University Av	375-3222	Abbott Lynn K 1365 N 1220 W	374-5382
A & Y BUILDING SUPPLY INC		Abbott Margaret A 525 N 650 E Orem	225-6588
590 S 100 W	374-6485	Abbott Myra J 155 S 1200 W Orem	226-3256
A & Y Components		Abbott Myron Lynn & Sharlyn	
1418 S Industrial Parkway	375-5256	2675 N 1060 E	377-2097
Aagard Rick 501 W 1560 S	375-7546	Abbott Nanci Tingey Hall	377-7759
Aanerud Elliott Roy 696 S 500 W	377-5146	ABBOTT PEARCE THORN & HILL	
Aanerud Fred 492 W 1300 S	373-9524	attys 2696 N University Av	375-8900
Aaron Building Systems Inc		Abbott Ralph 2918 Marcrest West	374-9028
1160 S State Orem	224-0483	Abbott Rick 695 N 300 E	375-1531
AARDVARK FILM PROCESSING		Abbott Ronald K & Sue 295 N 210 W Orem	224-4062
Gallery 28 University Mall Orem	226-7531	Abbott Scott W Hall	375-4419
Aaron Charles 974 W 600 S	377-4361	Abbott Thomas 2330 W 200 N	375-6252
Aaron Christopher 188 S 1920 W	373-1517	Abbott W Nelson 269 S 700 E Orem	224-8844
Aaron G William 633 S 700 W Orem	224-4847	Abdalmajeed Mansour Akmed	
AARON JOSEPH N phys 1029 N 500 W	373-5870	1200 Terrace Dr	375-9690
Aaron Publishing & Indexing		Abe Deborah & Arthur R 263 E 600 N	375-7155
329 N State Orem	226-1162	Abegg Joseph B 926 S Main Orem	225-2390
Aaron's Rooter Service 124 E 1850 N Orem	377-2985	Abegg Myrlon B 513 Kwanzan Cir Orem	225-4884
Aaron's Sewing Machine Repair Co		Abegglen Joseph H 1211 N 1220 W	377-3334
269 W 400 N	375-0389	Abegglen Patricia 763 E 820 N	374-9787
		Abel Edward E 956 E 1200 N Orem	224-2902
		Abel JoAnn 378 S 1350 E	374-6147
		Abel Kenneth E 378 S 1350 E	374-6147
		Abel Kent J 377 S 530 E Orem	225-5525
		Abel Ralph J & Dela 574 W 1020 S	373-5983

We used a segment of the first page of the 1982 Provo-Orem, Utah, phone directory to illustrate a systematic sample. There were an average of 320 names on each of the 107 pages for a total of 34,270 names (320×107). We wanted a sample of 1,000 individuals and thus the interval between names is 34 ($34,240/1,000 = 34.2$).

A number between 1 and 34 was randomly selected by numbering slips of paper 1 through 34, mixing them in a box, and selecting one. The number was 4 and, as noted above, the fourth name, Aaron, Charles, was selected and then every 34th name after that was included, the next one being Abel, JoAnn. This process continued to the end of the directory listing. Note that businesses were skipped over.

ples of each stratum are selected. Within each subgroup or stratum a particular sampling fraction is applied, but the sampling fraction for some strata may differ from those of others. Stratified samples can be used only when information is available to divide the population into subgroups.

A researcher studying a community with a small black population may be concerned that too few blacks would appear in a simple random sample. He or she could obtain two sampling frames, that is, separate lists of whites and blacks. For example, the researcher might use, or perhaps create, a directory of the black community in order to stratify the names in the city directory according to ethnic identity. The same percentage, say 12 percent, could be selected from each of the two frames. In this way the sample would contain the appropriate percentage of blacks and whites and the researcher could generalize the findings to the population with confidence.

Disproportionate Stratified Sample

A **disproportionate stratified sample** is used when the researcher wishes to compare a subgroup or strata to other subgroups or to the population as a whole. The researcher first identifies the strata and then oversamples the small strata of interest by selecting higher percentages of them than of the rest of population.

In the example above, 10 percent of the black population could be selected as compared to only 2 percent of the whites. Through oversampling of blacks, a sufficient number would be obtained to compare with whites. It is important to remember that, since blacks are overrepresented, the findings cannot be generalized to the population. A useful strategy would be to obtain a stratified sample to generalize to the population and then select a disproportionate sample of blacks to compare with whites.

Cluster Sample

A **cluster sample** involves the division of the universe to be studied into groups or clusters of units (individuals, households, families, etc.). Then clusters are randomly or systematically selected from the universe of clusters, and all units within each selected cluster are studied. The advantage of cluster sampling is efficiency in data collection. The fact that the subjects to be interviewed or observed occur in sets or groups reduces the researcher's travel cost and the time necessary to collect the data.

To select a cluster sample, the researcher first obtains or creates a sampling frame of clusters of units. For example a sample of students in a school district may be obtained by drawing a sample of classes and then studying the students in these classes. Thus the potential respondents are clustered together in a relatively few classes rather than being scattered throughout the school district.

Although cluster sampling does curtail expenses, it unfortunately increases potential bias between a sample and the population, because the relatively small number of clusters studied may not represent the universe as well as would a simple random sample. If a higher level of bias or

threshold of possible error is acceptable in a given study, or if the study has a limited budget, then cluster sampling may be the appropriate sampling technique.

Occasionally multiple cluster samples are used to obtain a sample from a large population. First a sample of large clusters of sampling units is selected. If the sample has two stages, then a sample of the units is drawn from the chosen clusters. If the sample has three stages, then a sample of smaller clusters is selected from the larger clusters. For the third stage a sample of units is drawn from the second-stage smaller clusters.

An example of a multistage cluster sample is the sampling of all students in American public schools for the study of the quality of education conducted by James Coleman and his associates (1966). They desired to select a random sample of students in public schools to determine their academic achievement and to ascertain the quality of facilities, programs, and teachers available to them. Time and cost prohibited drawing a simple random sample of students. The researchers instead used counties as clusters of students and drew a sample from the 3,130 counties in the United States. They could have used the nation's 24,446 school districts as clusters but opted for counties because demographic information such as race, education, income, family size, and marital status are available from the U.S. census for counties. In order to ensure a better representation of students in various settings, census information was used to stratify counties by population size, geographic region, and proportion of nonwhite population. A stratified cluster sample of 332 counties was selected.

The second stage of the sampling process was to use schools as clusters of students. A sample of high schools within the selected counties was drawn. A list of the 4,522 high schools in the selected counties was obtained and a sample of 1,170 was drawn. In addition, the elementary, middle, and junior high schools that fed students to the selected high schools were included in the study.

The third stage of clustering was inherent in the decision to use entire classes or grades as clusters of students. Rather than select a sample of grades, the 1st, 3rd, 6th, 9th, and 12th grades were selected, and all students in these grades in the sampled schools were potential respondents. Of the approximately 900,000 students in the final sample, data were collected from over 600,000, as well as from 67,000 teachers and 4,000 principals. The fact that the students, teachers, and principals were clustered in classrooms and schools in only 320 counties greatly facilitated the data collection.

Area Sample

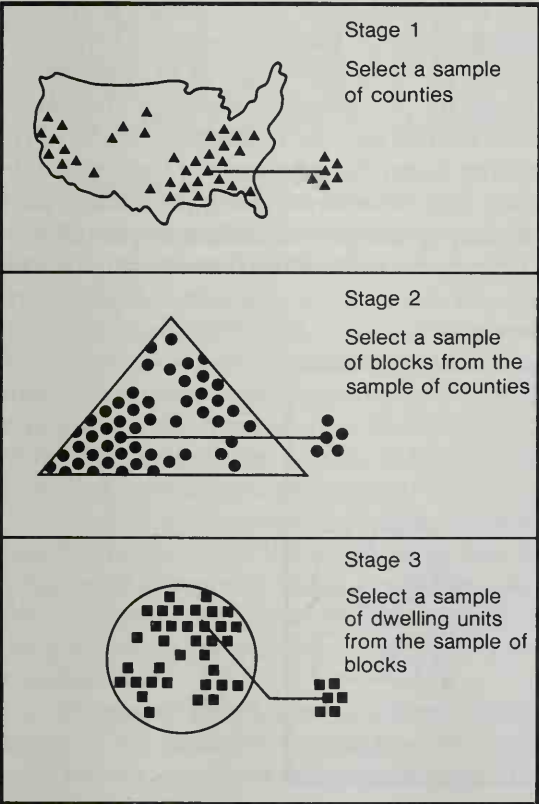
Area probability sampling is a special type of cluster sample and is generally used for national or other large populations for which adequate sampling frames are not available. As with the national study of the quality

of education discussed above, area probability samples are multistage. The first stage is to use primary geographical units as clusters of individual respondents. The clusters most often used are counties, census tracts, census enumeration districts, voting districts, and tax districts. A sampling frame of the geographical clusters is obtained or created from maps. A sample of the clusters is then chosen (see Illustration 3.3).

The second stage is to select a sample of smaller geographical clusters within the earlier chosen areas. Thus, if a sample of counties was selected, the second stage would be to draw a sample of census blocks from within the counties sampled.

The third stage would be drawing a sample of dwelling units in the blocks chosen in stage two. Often researchers may obtain from the Bureau of the Census, city planners, or county planners maps that identify dwelling units in census tracts, blocks, or other districts. Systematic procedures are developed for selecting dwelling units within the smallest geographic area in the sampling design. For example, a number between 1 and 4 can be selected by throwing a die to determine which corner to start at. A 1 could

ILLUSTRATION 3.3. Multistage area sample.

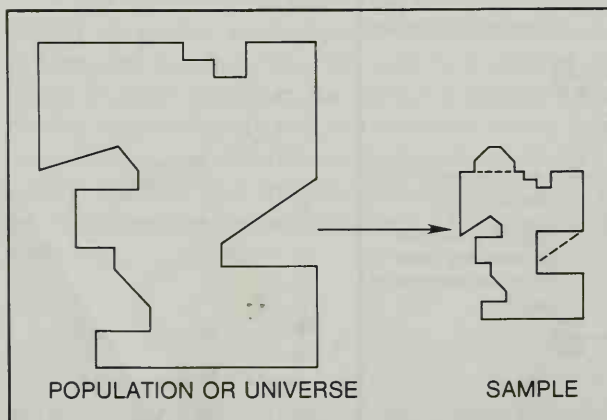


signify to start at the northeast corner, 2 at the northwest corner, and so on. A coin can be flipped to determine whether a researcher goes clockwise or counterclockwise in circling the selected block. The number of dwelling units on the block can be counted and then a sample selected. The critical element is that the process of selection be carefully described and be random (i.e., homes are selected by flipping a coin or selecting from a table of random numbers) rather than by some possibly biasing characteristic, such as whether there is someone at home when an interviewer first calls.

The number of stages varies from study to study, with national studies requiring a stage or two more than state or county studies. Regardless of how many stages are used, the principle is the same: selecting smaller and smaller clusters of sampling units until one arrives at the ultimate level where potential respondents are selected.

Whatever type of random sample is drawn, random selection will result in a sample that approximates the total population (see Illustration 3.4). The population in this illustration is drawn with an irregular shape to illustrate its unique characteristics. The similarity in shape between the population and the sample represents the accuracy of generalizations that can be made from the sample to the population. This accuracy is affected by the completeness of the sampling frame, the randomness of the sampling procedures, and sample size. These factors singly or in combination can produce a sample different from the population as shown in Illustration 3.4 by the bump on the top of the sample and the indentation on its right-hand side. An incomplete frame may mean that a segment of the population is not represented in the sample. Nonrandom sampling procedures may overrepresent some segments of the population and underrepresent others. Finally, a smaller sample has greater potential to differ from the population,

ILLUSTRATION 3.4. A sample of a population.



because one has fewer units of information from which to construct an image of the population.

III. NONRANDOM SAMPLING

Many texts on sampling do not even discuss **nonrandom sampling** because of its limited utility in scientific research (Kish, 1965; Sudman, 1967; Namboodiri, 1978). Although some types of nonrandom samples, such as convenience samples, are of little use to the researcher, snowball and quota samples may be useful at times.

Convenience Sample

A **convenience sample**, sometimes called a man-in-the-street sample, is just what the name implies: the study of sampling units that are conveniently available to the researcher. Results from convenience samples frequently appear on TV nightly news. Public reaction to the assassination of a political figure, a new defense system, a change in foreign policy, sporting contests, and similar events are "gauged" by the media using this type of sample. A roving reporter with a camera operator has interviewed individuals stopped on a busy street or in a shopping mall. Such interviews are interesting, but those interviewed do not represent any defined population and their comments are useless in ascertaining public sentiment.

Purposive Sample

For a **purposive**, or **judgment**, **sampling** the researcher uses his or her expertise to select subjects who represent the population being studied. The researcher chooses, on the basis of experience or other criteria, "average" blue-collar workers, "average" high school athletes, "average" teenage mothers, and so on. Although the researcher's trained judgment may produce a more representative sample than does simple convenience, there is no way to determine whether the sample in fact represents the population unless a representative sample of that population is studied and its characteristics compared with those of the purposive sample.

Despite this serious limitation, purposive samples are occasionally used by researchers. Voting districts that have a history of consistently voting as the majority voted are sometimes used as purposive samples. Voters in them are questioned about how they voted or intend to vote, and the results are generalized to a population such as the county, state, or nation.

Purposive samples are also occasionally used in studying infrequent behavior. For example, it would be difficult to find enough mass murders in a random sample of the general population to justify generalization of

any kind. Instead, the researcher interested in such behavior necessarily selects those few mass murderers or assassins who are known, who are accessible, or about whom sufficient biographical material exists to justify study (for an example of such purposive sampling in the study of mass murder, see Bolitho, 1926). Similarly, a purposive sample of well-known professional basketball players might be used to study aspects of the life of professional athletes. Thus, purposive samples are at times useful and may produce insightful qualitative work or good journalistic analysis, but the researcher and research consumer must remember that the "sample" described may not accurately represent a wider population.

Snowball Sample

Snowball sampling identifies a few research subjects who have characteristics relevant to the study and in the process of data collection asks them to name others they know who are like them in the relevant characteristics. The process is repeated until the researcher has obtained the desired number of research subjects. The name "snowball" aptly describes the process; a small sample grows larger as those identified name others. A snowball sample, like a purposive sample, may in fact produce a group of subjects who represent the population to which generalization is made, but the researcher does not know to what degree the sample represents the larger population. Therefore the findings necessarily remain illustrative.

Although this is a limitation, there are times when snowball sampling is the most efficient way to identify subjects relevant to a given study. For example, an environmental impact assessment of a large coal strip mine proposed on remote public lands would require the identification of concerned citizens, including environmentalists, sportspeople, businesspeople, and government officials who can provide information about likely consequences of mining in the proposed area. Snowball sampling has proven very efficient in accomplishing this task. Local political leaders responsible for managing the land may be interviewed and asked to identify other persons and interest groups who have relevant experience or opinions about the proposed project. Those named may in turn be questioned about their feelings and experiences and then asked to name others knowledgeable about or interested in the project. The snowballing may continue until the names that are given are those already contacted, at which point the researcher is confident that most if not all of the relevant "expert" informants have been identified.

Quota Sample

Quota sampling uses information about certain characteristics of the population to identify a sample; a predetermined number of interviews would be obtained (or observations made) from respondents who have

these “essential” characteristics, such as sex and age. For example, a very simple quota sample might be the first 25 women who agreed to be interviewed about a topic. Often more representative quota samples are selected by obtaining a sample with the same proportion of individuals with selected characteristics as exists in the population being studied. For example, if the population is 48 percent male and 52 percent female, the same percentages of men and women would be included in the quota sample. Age, geographic region, marital status, education, race, and income are the characteristics most often used to establish quotas.

Research assistants or interviewers are assigned certain quotas of people with specific characteristics. At first, most potential subjects will fit into one of the predetermined categories, but as the quotas for more typical respondents are filled, finding a person of certain age, marital status, and education may be quite difficult. Obviously, the accuracy of generalizations from a quota sample depends on the accuracy of the information about the population used to establish the quotas. Accurate and detailed information may permit the development of finely described quotas that are representative of the population.

Quota sampling is nonrandom and therefore has the potential for abuse and bias. First, if the information used to develop the quotas is inaccurate, the various strata are over- or underrepresented in the sample. Second, studies using quota samples generally rely on first calls to obtain their data. Rather than returning for a second call, the researcher usually locates someone else with presumably similar characteristics to fill the quota. People who are “available” have a much higher chance of being included in quota samples than are persons who are away from home a lot or who are difficult to contact for other reasons. “Available” and “unavailable” persons may be quite different and thus biased results are obtained because the “unavailable” are underrepresented.

Even though quota samples are nonrandom, if they are done systematically and attention is paid to the problems of availability and the potential biases it introduces, they may provide generalizable results. In fact, professional research firms like the Gallup, Harris, and Roper organizations sometimes use quota sampling successfully.

SAMPLE SIZE

Students frequently ask, “How big a sample do I need?” There is no simple answer to this question. The size of the sample required depends on the nature of the population, the purposes of the study, and the resources available.

The heterogeneity of the target population is very important in deciding sample size. The greater the heterogeneity, the larger the sample

required to represent the population. If the researcher is interested in assessing the frequency of cheating on examinations at a university and most students cheat, then a very small sample may suffice. On the other hand, if there is a great deal of variation in cheating, with some students cheating all the time, others regularly, others infrequently, and some never, then a larger sample is required to ensure that a representation of all "levels" of cheating is included. Generally, researchers should assume that considerable heterogeneity exists in the characteristics under study unless they have firm evidence to the contrary.

If sufficient information is available about the population, the degree of heterogeneity can be calculated; it is statistically expressed as the *standard deviation*. The heterogeneity (standard deviation) of the sample also can be computed and compared with the standard deviation of the population to determine the standard error, which is the best estimate of sampling error. The sampling error can be used to establish confidence levels, or assurance that a sample of a certain size represents the population within certain specified narrow limits. For example, it may be determined that a sample of 500 students would reveal the incidence of cheating within five percentage points either way. However, such statistical estimates require considerable advance information about the population to be studied, and such parameters are usually not available. The problem is compounded when the research project includes several variables, each with a different degree of heterogeneity in the target population. Therefore, rarely can the social scientist turn to statistical considerations for the answer to the question about how large the sample should be.

Instead, researchers have developed certain rules of thumb about sample size. According to Bailey (1982), 30 is considered by many as the minimum size of a sample. Others opt for a minimum sample of 100 units, and we encourage the selection of at least 200 cases. Social research generally requires analysis about the effects of background characteristics on the behavior being studied, and each additional characteristic of interest adds to the potential ideal size of the sample. For example, we may be interested in determining the difference in cheating between men and women students. The ability to make meaningful comparisons would require twice as many respondents as would analysis if gender were ignored. To further assess the difference in cheating between married and single students produces subsamples of smaller size. Now one needs enough of each gender to further divide them into single and married students. Academic major and year in school are two other relevant characteristics which, if taken into account, might divide the sample into even smaller subsamples. A fairly large sample, say 200, is required to allow simple percentage comparisons on more than two characteristics.

Two often important considerations in determining sample size are the time and money available to the researcher. If the results of the study

are needed quickly to make some policy decision, the research process can be speeded up by limiting the sample size. Also, researchers may be forced by financial constraints to use smaller samples than would otherwise be preferred.

In sum, there are no absolute rules that dictate the size of sample. The researcher must review the nature of the population, estimate the completeness of the sampling frame, consider the nature of the behavior to be studied as well as the time and funds available, and then make a decision about sample size.

V. SUMMARY

The process of social research is simplified because not everyone in a population must be studied before generalizations can be made about the population. Instead, samples may be drawn, and if they are drawn correctly, results from the study of samples will accurately represent the wider populations from which the samples were drawn. The process of sampling greatly reduces the time and money required to do social research. A sampling frame is a list of sampling units—individuals, families, households, business executives, students, and so on—in the population from which a sample is selected. City directories, telephone books, membership rolls, and utility company lists are possible sampling frames.

Random sampling means that every individual in the population has an equal chance of being selected for the sample. For an unrestricted sample, each unit selected is placed back to the sampling frame before another unit is drawn. This ensures that the probability of selection remains constant. An unrestricted sample is typically used in the study of small populations. A restricted sample does not replace selected units to the sampling frame and thus the probability of the first unit selected is lower than the probability of units selected later. However, with larger populations this difference in probability of selection between the first and last units drawn is trivial and can be ignored. A systematic sample is a specific type of restricted sample. Every n th unit in the sampling frame is selected, after a number between 1 and n is randomly chosen as the beginning point. Systematic samples are considerably quicker and cheaper to draw than other samples.

Stratified samples are selected to ensure that appropriate numbers of members from small subgroups are included in the sample. The population is stratified by creation of a sampling frame for each stratum and a sample of each is selected. The percentage drawn from each stratum may be the same, or may be different according to the needs of the project. If different selection ratios are used, the process is sometimes referred to as disproportionate stratified sampling.

A cluster sample is the selection of groups or clusters of sampling units. All of the units in the cluster are then studied. The clustering of respondents greatly reduces the cost of data collection, for each "visit" to an area or "entry" to an institution guarantees many potential respondents. An area probability sample is a multistaged cluster sample that starts with a sample of large clusters like counties and then shifts to smaller clusters such as towns, census tracts, or blocks.

Nonrandom sampling is discouraged, as it is risky to generalize from nonrandom samples to larger populations. This is especially true for convenience samples. Purposive samples, whereby the researcher chooses subjects presumed to represent the population, and snowball samples, whereby initial subjects identify others relevant to the study, are useful at times, although there is always uncertainty about how well they represent the wider population, if at all.

Quota samples, which use information about the population to determine how many units with specific characteristics are to be obtained, may provide reasonably accurate generalizations. Quota samples are most useful when sufficient information is known about the population to determine appropriate quotas on a variety of characteristics.

The size of sample required is decided on the basis of the nature of the population, the topic being studied, the completeness of the sampling frame, the resources available to the researcher, and the degree of accuracy necessary for the project to be worth doing at all.

REFERENCES

- ALBRECHT, STAN L., HOWARD M. BAHR, and KRISTIN GOODMAN, 1983. *Divorce and Remarriage*. Westport, Conn.: Greenwood Press.
- ALBRECHT, STAN L., and PHILLIP R. KUNZ, 1980. "The Decision to Divorce: A Social Exchange Perspective." *Journal of Divorce* 3, No. 4: 319-337.
- BAHR, HOWARD M., BRUCE A. CHADWICK, and JOSEPH STAUSS, 1972. "Discrimination Against Urban Indians in Seattle." *The Indian Historian* 5, No. 4 (Winter): 4-11.
- BAILEY, KENNETH D., 1982. *Methods of Social Research*, 2nd ed. New York: The Free Press.
- BOLITHO, WILLIAM, 1926. *Murder for Profit*. New York: Time, Inc. edition.
- CHADWICK, BRUCE A., and JOSEPH STAUSS, 1975. "The Assimilation of American Indians into Urban Society." *Human Organization* 34, No. 4 (Winter): 359-376.
- COLEMAN, JAMES S., and Associates, 1966. *Equality of Educational Opportunity*. Washington D.C.: U.S. Government Printing Office.
- GLASSER, GERALD J., and GALE D. METZGER, 1975. "National Estimates of Nonlisted Telephone Households and Their Characteristics." *Journal of Marketing Research* 12 (August): 359-361.

- GROVES, ROBERT M., and ROBERT L. KAHN, 1979. *Surveys by Telephone*. New York: Academic Press.
- KISH, LESLIE, 1965. *Survey Sampling*. New York: John Wiley.
- MENDENHALL, WILLIAM, LYMAN OTT, and RICHARD L. SCHEAFFER, 1971. *Elementary Survey Sampling*. Belmont, Cal.: Wadsworth.
- NAMBOODIRI, N. KRISHNAN (ed.), 1978. *Survey Sampling and Measurement*. New York: Academic Press.
- PERRY, JOSEPH B., JR., 1968-1969. "A Note on the Use of Telephone Directories as a Sample Source." *The Public Opinion Quarterly* 32 (Winter): 691-695.
- SUDMAN, SEYMOUR, 1973. "The Use of Telephone Directories for Survey Sampling." *Journal of Marketing* 10 (May): 204-207.

Observational Methods

four

I. Introduction

II. Advantages of Observational Methods

III. Problems with Observation

IV. Doing Observational Research

Structure in the research setting

Content of observation

Aids in recording observation

*Developing category systems
for recording observations*

Sampling behaviors for observation

Training observers for observational research

V. Analyzing Observational Data

VI. Illustrative Examples

*Unstructured observation: on
being sane in insane places*

*Structured observations in more
controlled settings*

*what does it mean to smile?
changing the teaching game*

VII. Summary

I. INTRODUCTION

This and the next several chapters deal with substantive methods for conducting social research. **Observation** is clearly the most basic method for obtaining information about the world around us. We all observe, and we tend to do it on a regular basis, although the degree to which we do it systematically varies from individual to individual and from setting to setting:

People-watching from the comfort of the front porch or the sidewalk cafe is a time-honored custom. The [social scientist] does not confine himself to a laboratory, a library, or to testing and interviewing. He often observes social behavior as it takes place in the natural environment. However, unlike the layman, the [social scientist] plans his observations and systematically selects setting, procedures, and measurements before venturing into the field. (Tedeschi and Lindskold, 1976: 61)

The principle of observation underlies all of the methods used by scientists in their data gathering. Sometimes the social researcher directly observes whatever he or she is studying. Other times the researcher relies on the reported experiences of others as they are recorded on questionnaires or in interviews. In still other cases the researcher uses "processed" materials, such as other studies, census reports, directories, and the like

as the main data source. In all instances, the principle of observation operates to some degree—the researcher observes personally, refers to the recorded observations of others, or asks people to tell about their own behavior and/or attitudes. In short, observation is the basis of laboratory experiments, field studies, participant observation, interviews, and the ultimate source of all secondary data.

Chapters 5 and 6 will discuss interviewing and questionnaire research, techniques used when subjects are asked to describe or report their own observations about their behavior and attitudes. Chapter 7 will focus on the use of experimentation in social research. Chapter 8, dealing with qualitative research, will give primary attention to participant observation, one of the principal applications of observational techniques, and Chapters 9 and 10 will discuss the use of observations made by others for research objectives not envisioned by the initial observer. As is obvious, all of these methods entail the important principle of *observation*. This chapter will focus on some of the more general techniques involved in observation and will give special attention to those procedures for *directly* observing objects, situations, or people. Applications of these procedures in a variety of other research contexts—all of which will receive more detailed attention in later chapters—will be identified and discussed.

II. ADVANTAGES OF OBSERVATIONAL METHODS

Observation offers some clear advantages over other methods. One of the advantages of direct observation of behavior is that it allows the researcher to record behavior as it occurs, as seen by a disinterested outsider, rather than relying on a subject's retrospective or anticipatory reports of personal behavior (Selltitz et al., 1959). It is a technique that is largely independent of a subject's ability or willingness to report his or her behavior. A specific example of this can be seen in the research on patronizing of adult bookstores and attending adult movie theaters.

Many patrons of adult bookstores and theaters apparently feel some embarrassment about their actions and so researchers have been unsuccessful in obtaining questionnaire and interview data from them. However, several studies have been done of patrons of these establishments using observational techniques. Massey (1970) reports unobtrusive observations of almost 2,500 people in Denver, Colorado, who entered two bookstores that carried sex-oriented materials. Trained observers attempted to classify each patron in the store according to a variety of sociodemographic characteristics including sex, age, ethnicity, type of dress, and whether he or she wore a wedding band. Similarly, Nawy (1970) observed 950 patrons of 11 adult bookstores in San Francisco, and Winick (1970) collected observational data on 1,800 customers in adult bookstores in New York, Los

Angeles, Chicago, Detroit, Atlanta, and Kansas City. Winick also observed 5,000 patrons of adult movie theaters in nine communities that varied considerably in terms of size and cultural, ethnic, and socioeconomic characteristics.

All of these studies report that patrons of such establishments have remarkably similar characteristics. The typical profile that emerges is that of a white, middle-aged, middle-class, married male who is dressed in neat attire. The male is usually alone and typically avoids contact or interaction with other patrons. If someone moves close or communicates, the patron tends to move away in order to maintain some degree of distance between himself and other patrons. Their general hesitancy in responding to others makes unobtrusive observation not only the most effective but perhaps the only technique for valid data collection.

The other major advantage of observational techniques lies in their broad application in the social sciences. Such breadth is evident from the early definition of observational methods given by Selltitz and her colleagues:

Observation may serve a variety of research purposes. It may be used in an exploratory fashion, to gain insights that will later be tested by other techniques; its purpose may be to gather supplementary data that may qualify or help to interpret findings obtained by other techniques; or it may be used as the primary method of data collection in studies designed to provide accurate descriptions of situations or to test causal hypotheses. Observation may take place in "real-life" situations or in a laboratory. Observational procedures may range from almost complete flexibility, guided only by the formulation of the problem to be studied and some general ideas about aspects of probable importance, to the use of detailed formal instruments developed in advance. The observer may himself participate actively in the group he is observing; he may be defined as a member of the group but keep his participation to a minimum; he may be defined as an observer who is not part of the group; or his presence may be unknown to some or all of the people he is observing. (Selltitz et al., 1959: 204)

Used effectively, observation is a key method for collecting reliable and valid data over a broad range of human behavior. As a research tool, observation must be systematic. It differs from the casual observations that most people engage in daily in three ways:

1. The scientific observer pays particular attention to those categories of action determined by the researcher's specific objective. Other behaviors are ignored or given only secondary attention.
2. The scientific observer is more interested than most casual observers in identifying what appear to be the important factors associated with a particular action. In other words, why does a behavior occur? What seem to be its causes?

3. Unlike people generally, scientific observers are almost always interested in recording what is being observed by making tabulations, counting responses, classifying reactions, and so on.

III. PROBLEMS WITH OBSERVATION

The *direct* observation of behavior by the researcher carries certain disadvantages as well as advantages. Among the most important problems associated with scientific observation are the following.

1. There are problems related to the sheer inadequacy of human sense organs. In observation, the basic tools for data collection are the sense organs, although mechanical devices such as video cameras and recorders are often used nowadays. To the degree that observers must rely on their senses, the inadequacies of personal observation must be recognized. For one thing, what we see and what we hear are influenced by our mental and physical states. Observer fatigue and boredom are critical factors in many studies. After observing something for a time, people may begin to miss fine details, count inaccurately, or overlook important changes in the nature of the interaction or other behavior being observed. More serious, however, is that because of the limits of the sense organs, observers may literally not see or hear what goes on, or may misinterpret what is observed because only part of the situation was visible or audible to them.
2. Selective perception is often a problem. Even the best trained researcher may produce biased data because of **selective perception**. People tend to "sense" certain phenomena more than others, and when we center our attention on one thing, we risk missing something else that may be happening. We tend to perceive those phenomena that carry meaning from our own point of reference. A dramatic event may distract attention from something that actually has more theoretical relevance to the research objective.

The problems of selective perception in the observation of noncontrolled situations are clearly evident in studies of crowd behavior. A simple example is the estimation of crowd size. Jacobs (1967) cites an instance from the 1968 presidential election campaign when candidate Richard Nixon stopped at a Milwaukee airport to address a crowd. Several different estimates of the crowd size were made. Republican party officials estimated that there were 12,000 people at the airport. Police estimates put the crowd size at 8,000. A local newspaper reported that 5,000 people were present. And a photograph taken from above showed that about 2,300 people attended the speech. While party officials may have had something to gain by overestimating crowd size, the variant estimates by the police and local

newspaper, in contrast to the "hard" evidence of the photograph, illustrate the range of error in recording so objective a "fact" as crowd size.

3. Our senses are poor instruments for making comparisons because they adjust to conditions. A characteristic that initially seems very important may fail to retain our attention as it becomes commonplace. A classic example is found in William Foot Whyte's *Street Corner Society* (1943), in which Whyte notes that he moved from the role of a nonparticipating observer of events to that of a nonobserving participant. He became so accustomed to living with the Norton Street gang that he virtually ceased, for a time, to observe and record.

The problems of adjustment to conditions as they affect the participant observer were summarized several years ago by Goode and Hatt:

. . . to the extent that [the observer] participates emotionally, he comes to lose the objectivity which is his single greatest asset. He reacts in anger instead of recording. He seeks prestige or ego satisfaction within the group, rather than observing this behavior in others. He sympathizes with tragedy and may not record its impact upon his fellow members. Moreover, as he learns the "correct" modes of behavior, he comes to take them so much for granted that they seem perfectly natural. As a consequence, he frequently will fail to note these details. They are so commonplace as not to seem worthy of any attention. (1952: 122)

4. Our senses do not operate independent of our past experiences. Consequently, both what we observe and the interpretations we attach to what is observed are influenced by what we have previously seen, heard, felt, and done. A person sensitized by past experience to certain types of exchanges will see those things to which he or she may be sensitized, while the same events or nuances may be completely missed by someone from a different background having a different set of sensitivities.

5. Finally, the very process of observation may influence the phenomenon that is being observed. This problem was discussed earlier when we described the subject matter of the social sciences as "reactive." That is, human subjects often behave differently because they are being observed.

Various procedures are used to diminish the potential impacts of these five problems. Sometimes multiple observers view the same phenomenon, but even they produce only a few different perspectives, and their unanimity, if it occurs, is no guarantee that they have observed accurately. Or, several observers may produce several different descriptions as in the example of crowd size.

A second strategy that we discussed earlier is triangulation. Under such a procedure, observational data are supplemented by data obtained by other means such as questionnaires or secondary data sources. Accuracy is sought in multiple methods rather than multiple observers.

ILLUSTRATION 4.1. Frequently those being observed behave differently *because* they are being observed.



"DO YOUR STUFF — YOU'RE
ON MICROSCOPE."

© Sidney Harris. From *Chicken Soup and other Medical Matters*; Los Altos, Calif.: William Kaufmann, Inc. Reprinted by permission.

IV. DOING OBSERVATIONAL RESEARCH

There are some issues that must be given special attention when a research problem requires the collection of data through the *direct* observation of the subjects. The researcher must consider the amount of structure that will be built into the research setting; a decision must be made regarding

the specific content of the observations; aids that might be used to facilitate what is being observed must be designed and/or planned; categories for recording observations must be developed; and the researcher must decide whether to attempt to observe all behaviors and subjects of interest or whether some type of sampling scheme will be used. In addition, specific attention must be given to training those who have the most important role in observational research—the observers.

Structure in the Research Setting

Probably the most critical issue faced in any observational task is how much structure the researcher builds into the design. Highly structured situations impose a set of deliberate restrictions on the observer. These may include restricting the range of behaviors to be observed, restricting observations to only certain participants in an interaction, observing participants only in particular settings or at particular times, fitting observations into clearly defined sets of categories, or using a standardized system to record observations.

Structure and control are most evident in experimental settings. In such contexts, the researcher has defined beforehand the dependent variables to be observed, how they are to be measured, when and how an independent variable is manipulated, how its effects are measured, how extraneous variables that might affect the dependent variable are to be controlled, and how the recording of observations is to be done (e.g., by using audiovisual devices).

Unstructured observations are more apt to be used in “natural” settings, although sometimes highly structured observation is possible in the field. Usually the researcher who uses unstructured observation is interested in observing behaviors or events over which he or she can exercise little control. With unstructured observation, antecedent events are much more difficult to manipulate, and so the researcher has difficulty attributing causal properties to any particular independent variable. Typically the range of behaviors to be observed is much broader and less clearly defined in unstructured than in experimental settings. Unstructured observation is generally one of the techniques used in participant observation. The advantages and disadvantages of this approach are covered in greater detail in Chapter 8.

Whatever the degree of structure imposed on the research setting, the observer must usually consider four questions (Selltiz et al., 1959: 205):

1. What exactly is it that the researcher wishes to observe?
2. How are the observations to be recorded?
3. What procedures will the researcher use to ensure the accuracy of the observations made?
4. What should be the relationship between the observer and the observed?

The answers to these questions virtually determine all the characteristics of the research design. Further, the more structured the research setting, the more likely the researcher is to have determined *beforehand* the answers to these questions.

Structured observations typically focus on a clearly defined aspect of behavior. For example, in a small group setting, the researcher might choose to focus on such questions as who gives direction in the group and who provides socioemotional support to other group members (Bales, 1970) and might choose to ignore other types of interaction in the group. "Since the situation and the problem are already specified, the observer is in a position to set up in advance the categories in terms of which he wishes to analyze the situation" (Selltiz et al., 1959: 223).

The fact that many things have been determined before the researcher approaches the research situation does not mean that changes cannot be made. Specific types of interaction may take on increased meaning during a period of observation and the general focus of a project may change. Typically, however, before beginning the observational process, the researcher has decided what is to be observed.

In structured observation the method of recording observations is also usually determined beforehand. Before going into the field, the researcher may have developed a series of rating scales or various recording devices. In such situations, recording may simply amount to filling in categories or counting and entering responses of a certain type. The complexity and difficulty of recording procedures are determined by the complexity of the behavior and the category schemes developed to systematize the observation of that behavior.

There are several ways to increase the confidence one has in the observations. One means of increasing their accuracy is the use of several observers. However, different observers are likely to agree on what has been observed only if the behavior to be observed has been well-defined, the categories clearly developed, and the observers adequately trained. The importance of appropriate training cannot be overemphasized.

The final question, regarding the relationship between the observer and the observed, applies to both structured and unstructured observation. The problems that surround the relationship between the participant observer and the group being observed (see Chapter 8) also apply to other types of observation. Issues of entry to the group, whether the observation is to be overt or covert, the observer's role, and the possible effects the observer has on the group and its behavior must be considered in structured observation.

There are no easy solutions to these questions. In some situations, as was described in the example of patronizing adult bookstores and theaters, it would be virtually impossible to obtain the data one needs if those being observed knew what was happening. Another classic example of

this is Humphreys' (1970) study of middle-class male homosexuality, which will be discussed in detail in Chapter 8. Had Humphreys' subjects been aware of what he was doing, he would probably not have been allowed to complete his observations. In other instances, as in the research of Oscar Lewis, awareness on the part of the subjects that they were being studied was essential and permission was sought and obtained (cf. Lewis, 1959, 1961).

The general consensus about the effects of an observer on group behavior is that there is a sizable initial effect but that in a relatively short time the observer becomes less obtrusive and the group behaves as it would if the observer were not present. Even so, researchers engaged in direct or participant observation are constrained to assess the effect of their presence on the group's behavior.

Content of Observation

The "content" of what one observes is largely determined by one's research objective. If one is interested in questions of status and authority in group settings, then specific attention is given to situations that help to identify an individual's place in a group's status hierarchy, such as occasions where one grants or receives esteem, recognition, or deference.

Observation studies typically focus on several important dimensions of content. First, attention might be given to the *setting* in which the interaction occurs. Kenen's (1982) study of social interaction in the laundromat showed that physical settings are characterized by factors that promote or impede certain types of interaction. Such "setting" factors include things like the placement of the washing machines in the laundromat, the location of chairs where patrons can sit, tables for folding one's laundry, and so on. We all recognize that some settings are more conducive than others to certain types of behavior. For example, funerals elicit fairly predictable responses in most people. Therefore, initial attention might be given to the physical properties of the research setting and to the social definitions associated with that setting and the cultural context in which it is imbedded.

A second dimension likely to receive attention in many observation studies is the overt *behavior* of the actors, that is, who does what, with whom, and with what consequences? Studies of behavior in public settings where participants typically remain strangers indicate that people often engage in personal space-preserving behaviors. In the public laundromat patrons typically select a washing machine that is at least one machine removed from other patrons (Kenen, 1982). If there are empty seats available, bus riders usually will not sit next to anyone, and they may place objects in an empty seat next to their own to discourage others who might sit there. Civil inattention or carefully ignoring the presence of others is used as a technique to inform strangers that we are not interested in interacting with them (Goffman, 1963; Cary, 1978).

Observation may also focus on patterns of *verbal communication*. Attention might be given to the content of communication, its interpretation, and its consequences. Roles of leadership and following might be identified in this fashion. Patterns of interaction might be seen as being independent of the content of the communication.

An observational study will not necessarily select any of these separate types of content. Often the researcher is interested in how various types of content occur together, or on the influences between them. The issue might be, how does physical or cultural setting affect behavior, or are verbal expressions consistent or inconsistent with overt behavior?

Aids in Recording Observation

The quality of observational data can frequently be improved by means of aids or extensions of the senses. The particular technique one uses in recording observations depends on the amount of structure that is built into the research setting and the goals of the project. Unstructured observations will typically be recorded in some form of **field log**. For example, the researcher may keep a rather detailed diary in which he or she regularly records observations for subsequent analysis. If the group being observed is aware of the study, then extensive diary notes may be made regularly and openly. However, if the group is unaware of the role of the researcher, then note-taking will be made more from recall at the end of the day, or at other times when the researcher can find temporary privacy.

In either case, the accuracy, richness, and consistency of the data might be improved by supplementing the descriptive, narrative field notes with **schedules** or recording schemes designed in advance. Such protocols serve to sensitize the researcher to salient items or activities and also may serve as convenient ways to summarize such information as the socio-demographic characteristics of the respondents (age, sex, ethnicity, etc.), or the apparent power and communication patterns in the group.

In more structured observational settings, researchers use a variety of recording devices, including audio recordings, video recordings, motion pictures, still photographs, and other **mechanical recording devices** which keep a running tally of certain events that are keyed in by the researcher. The importance of such devices is that the data analysis may occur long after the events of interest have happened. The recordings then become a basic data source.

Recording devices can be used either **obtrusively** or **unobtrusively**. If they are used obtrusively, the researcher obtains permission from the subject to record an interview or to film an interaction sequence (for example, to film interaction between mothers and their children for later analysis of interaction patterns). If the devices are used unobtrusively, the subjects are unaware that recordings are being made of their behaviors.

Two of the authors of the text used the latter approach to study several large protest demonstrations in the late 1960s. Observers equipped with hidden tape recorders distributed themselves among a crowd of demonstrators. The tape recorders were used both to record statements made by members of the crowd and to record selective observations by the researchers. For example, as changes occurred in the nature of the crowd's activities, the observers would describe the changes, thereby entering them into the ongoing tape recording of the protest. Estimates of variations in crowd size were recorded the same way.

Developing Category Systems for Recording Observations

A critical part of systematic observation is the creation of an appropriate category (coding or conceptual) system for recording observations. A category is "a statement describing a given class of phenomena into which observed behavior may be coded" (Heyns and Zander, 1966: 389). A well-developed category system creates a shared system of viewing behaviors and reduces the chance that different observers will record the same activity in different ways.

The nature of a good category system depends in part upon the particular objectives of the researcher. However, several characteristics of good category systems are widely applicable. These characteristics include exhaustiveness, compatibility to inference, and number of dimensions (Heyns and Zander, 1966: 389-390).

Exhaustiveness Some category systems may be so universal that all possible behaviors might fit in one of the categories. Others are less exhaustive and are designed to treat some behaviors and not to apply to others. Researchers using a partial or middle-range category scheme would observe and record only behaviors clearly specified as belonging within the classification. All other behaviors would be ignored as far as the project at hand was considered.

Compatibility to inference Some observation is guided by coding schemes that require observers to make decisions about the meanings of the behavior observed, or to infer from the known and observable to the unknown and more abstract concept being studied. Other types of observation do not require coders or observers to make any inferences; working with such category systems requires careful observation but no inference or attribution of meaning, the latter being reserved for the analysts who later interpret the raw or objective observations mechanically collected by observers or electronic equipment. For example, an observer interested in interaction patterns between dating couples might, under a system requiring no inferences, record such behaviors as the following: "hold hands,"

"gently touch," "look into each other's eyes," and so on. Under a system requiring interpretation or inference, the researcher might use such categories as "shows affection," or "shows anger." In the latter case, the actions of holding hands, touching, and gazing at each other might all be considered acts of affection.

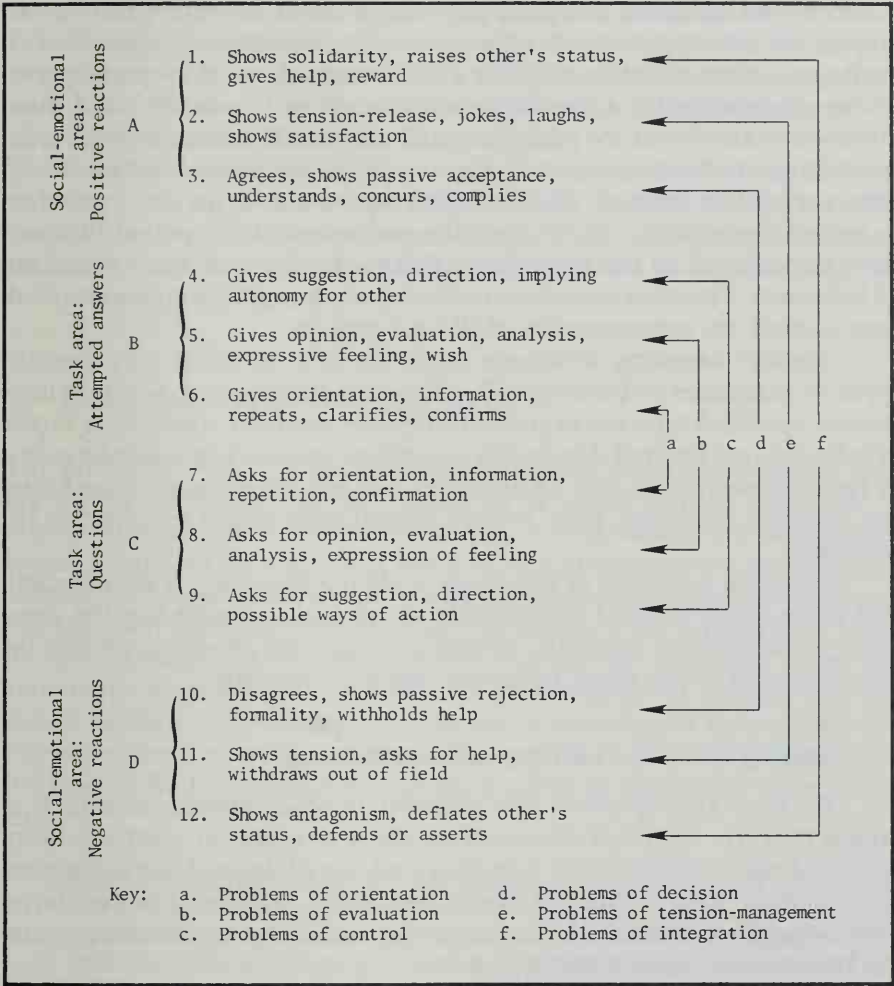
At some point in the research process, multi-category systems which make fine distinctions are usually collapsed into fewer categories to make analysis simpler. Sometimes such collapsing of categories is done by the observer; more frequently it is the task of the analyst. Of course, in many instances the observer and the analyst are the same person. Often the best procedure is simply to record behaviors in as fine detail as possible, leaving the collapsing or recombination of categories to the analysts. If the observers have to do too much decision making, their efficiency of observation and recording is likely to suffer.

The decision about who will do what is usually based on such considerations as the sophistication of the observer and the difficulty of the observational task. Usually making reliable inferences about psychological states and motivations for behavior is much more difficult than is the simple recording of behavior into predefined, well-bounded categories.

Number of dimensions Some category systems are developed to highlight a single dimension of behavior; other are more complex. For example, the discussion of dating behaviors above refers to a single, overt behavioral dimension. In other circumstances the researcher might be interested in attempting to assess subjective states as well as overt acts. Or, the researcher might be concerned with both the affective dimension of behavior and the intellectual, problem-solving activities of a group. As one moves from single to multiple dimensions, the tasks of observation and classification become much more difficult.

One of the most detailed, fully developed classification schemes for observational research was designed by Robert Bales (Bales, 1950, 1970). Bales' system provides a set of categories for classifying patterns of interaction in group settings (see Illustration 4.2). Using Bales' procedure, the observer records behaviors act-by-act as they occur. His twelve categories are assumed to be exhaustive in that all types of interaction that occur in a group setting can be coded into one of the types. Moreover, the content of each of the categories is described clearly and relatively little inference by the observer is required. An exception is that observers using Bales' system must make decisions about whether a given act "shows solidarity" or "shows tension." Finally, two basic dimensions of behavior are tapped by the system: goal-directed and socioemotional behavior. Goal-directed behaviors include activities like "gives suggestion" or "gives orientation." Socioemotional behaviors include tension release actions such as joking and laughing. The Bales method has been used extensively in research on

ILLUSTRATION 4.2. The original system of categories used by Bales in observation and their relation to major frames of reference.



Robert Bales, *Interaction Process Analysis*: by permission of The University of Chicago Press. 1950. © 1950 The University of Chicago Press.

small groups conducted over the last three decades. It is an excellent example of a well-defined category system for recording observations.

Sampling Behaviors for Observation

Frequently the researcher will have neither the time nor the resources to record all behaviors he or she is interested in or to record behaviors continuously. In such cases, a typical procedure is to sample. As is true

with other sampling procedures, one is generally interested in selecting samples of behaviors that would be representative of all behaviors that occur. Several different sampling procedures might be employed. For example, the researcher might sample according to time sequence. Under such a procedure attention might be directed toward one of the participants under observation for a specified period (such as 15 minutes) and then switched to another of the participants. During the 15-minute periods only the behavior of the primary respondent would be considered and recorded; others would be ignored. Another approach would be for the researcher to record observations about all of the participants for a period of time, then take a break in the recording activities, then return again to record all behaviors. The latter procedure rests on the assumption that the sampled time periods are representative of all time periods.

Another sampling technique might involve sampling only specific types or categories of behavior. That is, when behaviors of a certain previously specified type are expressed, they are recorded while other types of behaviors are ignored. Under this procedure, the decision might be made to record observations only when certain key behaviors have entered into the findings. Behaviors prior to their introduction would not need to be recorded.

Although sampling of behaviors is often a necessity, it is important that some clearly defined procedure be developed to ensure that the sampling procedures are adequate. Otherwise, what one obtains might not be representative of the actual behaviors that are occurring.

Training Observers for Observational Research

We have already noted that reliability in observational research is a critical concern. Different observers are likely to agree on what has been observed only if the behavior to be observed is well-defined, the categories for recording observations are clearly developed, and the observers have been adequately trained. We conclude this section by emphasizing again the importance of this training process.

The steps that make up a well-developed training program have been outlined by Selltiz et al. (1959: 232). First, the observers are given information about the purposes for the study. Its theoretical rationale might be explained and detailed attention is given to the logic behind the development of the categories that will be used to summarize and report the observations that are made. Second, the trainees are given the opportunity to observe behaviors of the type that will be the focus of the research project and to record their observations on a prepared schedule or data-recording sheet. Typically, the trainee will experience difficulty keeping abreast, making decisions on marginal cases, and selecting proper categories for recording observations. These difficulties are resolved through

discussion and further practice. Third, the trainees will participate in a pilot study that is designed to replicate on a reduced scale the actual experiences they will have in the field. Again, questions and problems are resolved. Fourth, groups of trainees are now ready for reliability checks. Several practice sessions, each followed by a reliability check, may be necessary until all of the observers become comparable measuring instruments. Finally, the trainees are ready for data collection.

In practice, most researchers do not give this much attention to training observers. However, the quality of the data obtained is largely a function of how good the observers are. If one cannot demonstrate reliability on this point, then little credibility can be granted to the data and the interpretations that flow from it.

V. ANALYZING OBSERVATIONAL DATA

The problems and difficulties associated with analyzing observational data are largely determined by the amount of structure that the researcher is able to build into the observational setting. If observations are made under highly controlled settings, such as experimental laboratories, and if observations are recorded within previously developed and clearly defined categories, then the analytical procedures are basically the same as those used to analyze other quantitative data (see Chapter 13). However, if the observational setting is unstructured and the categories of behavior are not clearly defined or are difficult to measure in quantitative terms, then different problems of analysis confront the researcher. This is well-illustrated in Howard Becker's description of the analysis of participant observation data.

Observational research produces an immense amount of detailed description; our files contain approximately five thousand single-spaced pages of such material. Faced with such a quantity of "rich" but varied data, the researcher faces the problem of how to analyze it systematically and then to present his conclusions so as to convince other scientists of their validity. Participant observation (indeed, qualitative analysis generally) has not done well with this problem, and the full weight of evidence for conclusions and the processes by which they were reached are usually not presented, so that the reader finds it difficult to make his own assessment of them and must rely on his faith in the researcher. (Becker, 1970: 190)

Becker proposes three basic steps in the analysis of observational data: First, the researcher selects and defines the problems, concepts, and indices to be used. In unstructured observational research these often will emerge while the researcher is in the field. In more highly structured research settings they will usually be specified in advance. In either case, the at-

tention of the analyst is directed to the problems and concepts that promise the greatest understanding of the group or organization being studied.

Second, the researcher measures the frequency and distribution of the events or characteristics being studied. For example, are events that prompt the occurrence of other events typical and widespread? In what way are these events distributed among the group? At this point the researcher is interested in quantitative estimates of how frequently certain events occur among the study groups and the context of the events. In unstructured observations these "counts" are typically made from one's detailed field notes after one has finished a period of observation.

The final stage of analysis entails incorporating the findings into a general descriptive model of the group that is being studied. This model then serves as the basic guide for the presentation and interpretation of findings.

The important point to emphasize in this brief description of the analysis of observational data is that the procedures of analysis depend on the amount of structure imposed on the research setting. The observer who has a well-defined problem before beginning the observation, who already has a well-defined system of categories, and who is looking for evidence to test specific hypotheses will analyze the data very differently from the researcher who has not identified specific hypotheses or organizing principles when the data collection is initiated. In general, the more quantitative one's observations, the easier they are to analyze and interpret. Qualitative data often are interesting from the human interest or illustrative point of view, but documenting that a set of observations are generalizable to any other setting is difficult, as is demonstrating that the patterns observed are even typical of the group or society in which the observation took place.

VI. ILLUSTRATIVE EXAMPLES

The issues we have reviewed on observational analysis can perhaps best be illustrated at this point by describing some actual projects. We begin with a relatively unstructured observational study conducted in several psychiatric hospitals. Then we turn to two studies conducted under more structured conditions. In the first of these, the researchers observed "naturally occurring" behavior and did not try to influence or alter that behavior. The second presents a direct test of some important research hypotheses and provides an example of the use of observation in a setting where the researchers manipulate independent variables and observe the consequences for the dependent variable.

Unstructured Observation: On Being Sane in Insane Places

A good example of the use of unstructured observation is D. L. Rosenhan's (1973) study of psychiatric hospitals. The study involves a classic role reversal: psychiatric patients (actually pseudopatients) observed and recorded the behavior of hospital staff, including psychiatrists, nurses, and attendants.

Rosenhan (a professor of psychology and law) and several collaborators, including a psychology graduate student, three psychologists, a pediatrician, a psychiatrist, a painter, and a housewife, were admitted to twelve different psychiatric hospitals. The researchers managed to be admitted by calling hospital admission offices and complaining of "hearing voices." The author describes the procedure: "Beyond alleging the symptoms and falsifying name, vocation, and employment, no further alterations of person, history, or circumstances were made." All of the pseudopatients were admitted to the hospitals following the interviews and, immediately upon admission, stopped simulating *any* symptoms of abnormality. In other words, each behaved "normally," except that the "patients" carefully observed how they were treated, even to the point of taking extensive notes. At first the note-taking was done secretly, but after it was determined that no one seemed to care if patients kept journals or not, the researchers were able to record their observations openly.

No one on any of the hospital staffs was aware of the study. All of the "patients" behaved normally while in the hospitals, but they were never detected. All except one were admitted with a diagnosis of schizophrenia, and all were eventually discharged with a diagnosis of schizophrenia "in remission." The length of stay in the hospital ranged from 7 to 52 days, with an average of 19 days.

Although no hospital staff personnel were suspicious of the pseudopatients, other patients were. Some fellow-patients voiced their suspicions to the researchers vigorously, but since the hospital staffs and not other patients were the focus of the observations, this created no major problems.

One of the primary behaviors the researchers were interested in observing was the process of "labeling." Rosenhan notes:

Beyond the tendency to call the healthy sick—a tendency that accounts better for diagnostic behavior on admission than it does for such behavior after a lengthy period of exposure—the data speak to the massive role of labeling in psychiatric assessment. Having once been labeled schizophrenic, there is nothing the pseudopatient can do to overcome the tag. The tag profoundly colors others' perceptions of him and his behavior. (1973: 252–253)

This observation was reinforced by the tendency on the part of the hospital staff members to interpret and even distort patients' behaviors in such a

way that they would be consistent with popular theories of the dynamics of schizophrenia. The assumption seemed to be that because the patient was in the hospital, he or she must be psychologically disturbed and, given a diagnosis of psychological disturbance, practically any behavior could be interpreted as a manifestation of that disturbance.

Other behaviors observed and recorded by the researchers included the amount of time that hospital staff actually spent on the floor with patients. The average amount of their total shift time that attendants spent on the floors was 11.3 percent, which included not only the time spent mingling with patients but the time required for doing such chores as folding laundry, supervising patients, directing cleanup, and so on. Nurses, physicians, and psychiatrists were even less evident, with the latter seldom seen on the wards and then virtually only when they arrived or departed.

To determine the uniqueness of the patient-professional relationship in the psychiatric hospital setting, the researchers carefully recorded the responses made by the staff to patient-initiated contact. The pseudopatients would approach a staff member with a request such as "Pardon me (Mr., or Dr., or Mrs. X), could you tell me when I will be eligible for grounds privileges?" The behavior exhibited by the pseudopatient during the request was never bizarre nor disruptive. In Illustration 4-3 data on responses to such requests are compared with data on responses to similar questions the researchers obtained in different institutional settings. Columns 1 and 2 summarize responses of psychiatrists, nurses, and attendants in the psychiatric hospitals. Column 3 summarizes responses to a request for infor-

ILLUSTRATION 4.3. Self-initiated contact by pseudopatients with psychiatrists, nurses, and attendants in psychiatric hospitals compared with similar contacts in university settings.

NATURE OF RESPONSE TO CONTACT	SETTING			
	Psychiatric Hospitals		University Campus	University Medical Center
	Psychiatrists	Nurses and Attendants	Faculty	Physicians
Moves on, head averted	71	88	0	0
Makes eye contact	23	10	0	3.7
Pauses and chats	2	2	0	7
Stops and talks	4	0.5	100	29.3

Source: Adapted from D. L. Rosenhan, "On Being Sane in Insane Places," *Science*, January 19, 1973.

mation made of faculty members on a university campus (the questions asked on the campus were, of course, different from those asked in the psychiatric ward but they did involve a request for comparable information). Column 4 summarizes responses to questions asked of physicians at a university medical center. In this case, the researcher would approach a physician with a request for information similar to the queries made of faculty members on the campus. In each of the university settings, several different questions were asked, creating a situation in which a busy professional person was asked to take time to help a stranger.

As can be seen, in the nonpsychiatric hospital settings, a busy professional was willing to pause and respond to questions asked by the researcher. In the psychiatric setting, most staff personnel would not even make eye contact with the researcher.

Rosenhan's study produced both "hard" and anecdotal data about the depersonalization and powerlessness felt by patients admitted to a psychiatric hospital. The richness of the data set is largely a result of the observational techniques used by the researchers. Other procedures such as experiments and questionnaire surveys might have been much less successful in allowing the researchers to understand what it is like to be a psychiatric patient. The role of the researcher required concealment, but at the same time it provided access to information that would not otherwise have been available to outsiders.

Structured Observations in More Controlled Settings

The use of more structured approaches in observational research can be demonstrated in a brief review of two very different observational studies conducted in experimental laboratory settings. Both studies used settings that parallel the everyday experiences of many people, namely clinics and schoolrooms. The first, however, included very little effort on the part of the researchers to control or manipulate the behavior of the subjects, whereas the second project sought to exercise control and manipulate change.

What does it mean to smile? Daphne Bugental has studied communication in families for several years. As part of this research she and some colleagues (Bugental, Love, and Gianetto, 1971) examined the connotative meaning of smiles in parental communication. Previous work on family communication patterns led them to hypothesize that in the homes children respond differently to smiles from fathers and mothers. Specifically, it was proposed that smiles are related more closely to positive evaluations such as friendliness, approval, or consideration if the smiler is male instead of female.

This hypothesis was based on the observation that women, and mothers in particular, may use smiles as part of their culturally approved role.

That is, the traditional female role demands more warm and compliant behavior, and smiling is a component of warmth and compliance. Thus, there may be little connection between a mother's smile and the content of what she is saying. For the mother, then, the smile may be role-defined rather than being directly related to the content of the immediate verbal exchange. For men, on the other hand, smiles may be used more often with friendly, approving statements. That is, they may be closely tied to the content of what a man says because approving behavior in the family context is less clearly prescribed for men.

To test their hypothesis, the researchers videotaped the interactions of 40 families and analyzed the videotapes for facial expressions and verbal content of communications. The videotaping was done in a room adjoining a reception area where the family waited for an appointment at a clinic. Each family was videotaped for about 20 minutes in the waiting room and also during a 5-minute session while they discussed the things they would like to change in their family. Following the taping, each videotape was evaluated for visual content by being shown to judges with the sound turned off. Typescripts of all parent-child messages in which the speaking parent was visible were judged independently for content by another group of judges. Visual ratings were then compared with verbal ratings.

In this project, the coding of observations was done from the *tapes* rather than the live sessions. Judges had clearly defined categories for sorting and clustering their observations. For example, visual ratings were divided among three categories: positive (smiling), neutral, and negative (negative facial gestures or expressions). Verbal content was evaluated on a five-point scale from positive (friendly, approving, or considerate) to negative (unfriendly, disapproving, or inconsiderate). Parent-child transcripts were retained for further analysis only if there was high agreement among the judges as to the meaning of the conversations.

Consistent with the hypothesis, the researchers found that mothers' verbal messages were no more positive when they were smiling than when they were not. On the other hand, fathers were much more likely to be making positive statements when they were smiling. The authors concluded that when fathers smiled at their children, they were sending a message that was more friendly or more approving than if they were not smiling. However, when mothers smiled, their verbal message was no more positive than if they were not smiling.

The procedures used by these researchers were clear and straightforward. Their hypothesis specified the specific behaviors they wanted to observe and how these were to be defined, and their procedures narrowed the focus of observation and coding to the data required for testing the hypothesis.

Changing the teaching game A classic example of the use of observation in a controlled experimental setting is found in the work of Robert Hamblin and several of his students (Hamblin et al., 1969). Hamblin and his co-workers designed a simple system for reinforcing "good" behavior in a classroom setting in order to change the behaviors of a group of extraordinarily aggressive boys. Specifically, the study was designed such that when a boy exhibited an "approved" behavior, he was rewarded with something he valued. At the same time, it was intended that activities by teachers or others that encouraged "undesirable" behaviors would be curtailed. The effectiveness of the proposed reinforcement program was evaluated via a simple token-exchange system. Approved behaviors were rewarded with a small plastic token. At the end of a specified period, the tokens could be exchanged for something the student wanted, such as a treat, a movie, or a walk outside the school. Throughout the study period the "aggressive" child was allowed to earn, then spend, earn, then spend, and so on. Because initially the tokens were meaningless to the child, they were paired with other reinforcers. For example, when a boy did an approved activity, he was rewarded with a verbal "thank you," a piece of candy, and a token. Eventually, when he had learned that the tokens could be exchanged for toys or privileges, the candy was withheld and the child was motivated by the tokens and the verbal reinforcements.

The setting for Hamblin's work was an experimental classroom at Washington University in St. Louis. Into this classroom Hamblin and his co-workers brought a group of young preschool boys with severe behavior problems (each child was extremely aggressive and had failed to respond to the therapy programs used in local schools). The teacher was told about the boys and about the general nature of the experiment and then sent into the classroom. She was instructed first to use whatever prior experience and training she had in managing the class. No tokens were used in this part of the experiment.

The observational procedures were highly structured. Trained graduate assistants watched the behaviors of each boy from an observation room behind a one-way glass wall. All behaviors were recorded for previously specified time periods. Behaviors of different types were coded, with primary attention given to those actions that could be defined as "aggressive," "cooperative," or "academic." The latter included any time the students spent listening to the teacher, looking at books, drawing, coloring, and so on. An observer was also assigned to watch the teacher and record her activities and responses to the students.

The teacher tried virtually every technique she knew: she was, on various occasions, a strict disciplinarian, a wise counselor, and a sweet peacemaker, but nothing worked. By the eighth day of the study the children were averaging about 150 aggressive acts per day. To illustrate their

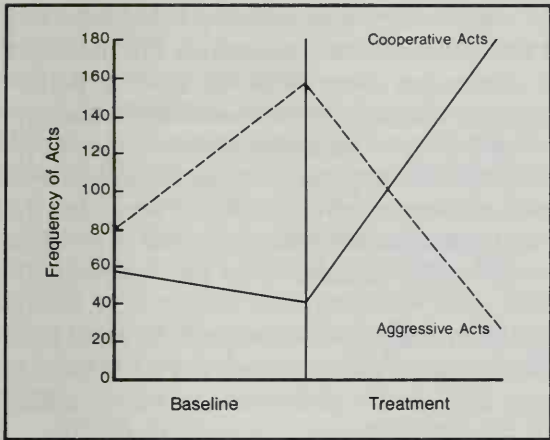
behavior and to provide an idea of what the observers saw and recorded, here is a four-minute sequence from one of those first days:

Mike, John, and Dan are seated together playing with pieces of Playdoh. Barry, some distance from the others, is seated and also is playing with Playdoh. The children, except Barry, are talking about what they are making. Time is 9:10 A.M. Miss Sally, the teacher, turns toward the children and says, "It's time for a lesson. Put your Playdoh away." Mike says, "Not me." John says, "Not me." Dan says, "Not me." Miss Sally moves toward Mike. Mike throws some Playdoh in Miss Sally's face. Miss Sally jerks back, then moves forward rapidly and snatches Playdoh from Mike. Puts Playdoh in her pocket. Mike screams for Playdoh, says he wants to play with it. Mike moves toward Miss Sally and attempts to snatch the Playdoh from Miss Sally's pocket. Miss Sally pushes him away. Mike kicks Miss Sally on the leg. Kicks her again, and demands the return of his Playdoh. Kicks Miss Sally again. Picks up a small steel chair and throws it at Miss Sally. Miss Sally jumps out of the way. Mike picks up another chair and throws it more violently. Miss Sally cannot move in time. Chair strikes her foot. Miss Sally pushes Mike down on the floor. Mike starts up. Pulls over one chair. Now another, another. Stops a moment. Miss Sally is picking up chairs, Mike looks at Miss Sally. Miss Sally moves toward Mike. Mike runs away.

John wants his Playdoh. Miss Sally says "No." He joins Mike in pulling over chairs and attempts to grab Playdoh from Miss Sally's pocket. Miss Sally pushes him away roughly. John is screaming that he wants to play with his Playdoh. Moves toward phonograph. Pulls it off the table; lets it crash onto the floor. Mike has his coat on. Says he is going home. Miss Sally asks Dan to bolt the door. Dan gets to the door at the same time as Mike. Mike hits Dan in his face. Dan's nose is bleeding. Miss Sally walks over to Dan, turns to the others, and says that she is taking Dan to the washroom and that while she is away, they may play with the Playdoh. Returns Playdoh from pocket to Mike and John. Time: 9:14 A.M. (Hamblin et al., 1969: 22)

Miss Sally always loses in such exchanges and often ends up, as above, reinforcing the very behavior that she was hoping to change. At this point the basic token-exchange system was introduced. Miss Sally was instructed to ignore negative behaviors but to reward all behaviors that were defined as positive (cooperation and studying behavior). Whenever a child did what the teacher desired, she was to give him a "thank you," an M&M candy, and a token. As noted, the tokens were small colored disks that could be exchanged at the end of the lesson period for various rewards. At first the children seemed motivated more by the M&M candies but soon they realized that the tokens would buy more desirable things. Eventually, the children were reinforced only verbally and with tokens. Data summarized in Illustration 4-4 demonstrate some of the most important changes in the children's behavior under the token reinforcement system. The average number of aggressive acts dropped from over 150 to under 60 per child per day. Similarly, the number of cooperative acts increased from 56 to about 115 per child per day. Later adaptations of the

ILLUSTRATION 4.4. Depiction of changes in aggressive and cooperative acts when a system of reinforcers is used.



Source: Adapted From Robert L. Hamblin et al., "Changing the Game from 'Get the Teacher' to 'Learn,'" *Transaction* 6 (January): 20–36, 1969. Reprinted with permission. © 1969 by Transaction, Inc.

program resulted in aggressive acts leveling off at about 16 per day while cooperation increased to about 140 sequences. Perhaps even more interesting is the time spent in lesson time. Before introduction of the token system the students paid attention to their teacher only about 8 percent of the lesson time. This figure rose to a plateau of 93 percent during the later stages of the token-exchange period, clearly a major shift in behavior.

The marked difference in the boys' classroom behavior during the token-exchange period as contrasted with the initial "baseline" period described earlier is apparent in the following sequence:

All of the children are sitting around the table drinking their milk; John, as usual, has finished first. Takes his plastic mug and returns it to the table. Miss Martha, the assistant teacher, gives him a token. John goes to the cupboard, takes out his mat, spreads it out by the blackboard, and lies down. Miss Martha gives him a token. Meanwhile, Mike, Barry, and Jack have spread their mats on the carpet. Dan is lying on the carpet itself since he hasn't a mat. Each of them gets a token. Mike asks if he can sleep by the wall. Miss Sally says "Yes." John asks if he can put out the light. Miss Sally says to wait until Barry has his mat spread properly. Dan asks Mike if he can share his mat with him. Mike says "No." Dan then asks Jack. Jack says, "Yes," but before he can move over, Mike says "Yes." Dan joins Mike. Both Jack and Mike get tokens. Mike and Jack get up to put their tokens in their cans. Return to their mats. Miss Sally asks John to put out the light. John does so. Miss Martha gives him a token. All quiet now. Four minutes later—all quiet. Quiet still, three minutes later. Time: 10:23 A.M. Rest period ends. (Hamblin et al., 1969: 24)

This study illustrates the use of observational procedures under highly controlled conditions. The researchers knew what dependent variables they wanted to observe: aggression, cooperation, and studying behavior. They were also able to control the introduction and variation in the nature of the experimental variable, the reinforcing token-exchange system. Behaviors were carefully measured and monitored in order to establish a baseline on each of the dependent variables prior to the introduction of the experimental variable. Changes in each of the dependent variables could then be accurately measured. Other conditions were controlled or remained the same under both conditions, and consequently changes in the dependent variable could be attributed with some confidence to the experimental variable.

The conditions of observation in this project were quite different from those in the previously described studies. Taken together, the three projects illustrate the range of problems that can be addressed via formal observation.

VII. SUMMARY

The most widely used method for obtaining information about the world around us is observation. We all observe constantly, and both receive and process the information obtained through our sense organs. For the social scientist, observation is essential to all methods of empirical research. The testing and development of a science of human behavior require the systematic collection and interpretation of valid and reliable observations.

Observation as a formal method of data collection offers some important advantages. First, the researcher records behavior as it occurs rather than having to rely on the retrospective or anticipatory reports of others. Second, observational methods can be used in many settings, from natural, everyday behavior to highly controlled laboratory experiments. Finally, observational methods allow the collection of data that might not otherwise be available to the researcher because of the inability or unwillingness of subjects to talk about themselves.

Problems associated with observation typically include those associated with the inadequacy of human sense organs; selective perception or the tendency to pay attention to certain events at the expense of others that might be as theoretically interesting; problems associated with the fact that observers adjust to conditions and become less effective recording "devices" because of overfamiliarity, boredom, or fatigue; the inability to observe and interpret independent of past experiences; and the potential effect of the observational process itself on what is being observed.

Most observational studies focus on one or more of three basic dimensions: the setting in which the interaction occurs, the overt behaviors

of the subjects in the setting, or the patterns of communication among subjects. The quality of observations can often be enhanced by the use of mechanical or electronic recording devices. Such devices might range from extensive field diaries kept by individual researchers to highly sophisticated audio and visual recorders.

Observational methods generally require the development of a system of categories for summarizing observations. Such a system provides a frame of reference designed to enhance the probability that all relevant aspects of the subjects' behavior will be noted. The important dimensions of a category system include exhaustiveness, inference, and the number of dimensions to be tapped by research. It is impossible to observe all behaviors over an extended period of time, and so sampling techniques are often used in observational studies. A sampling procedure might select certain time periods, subjects, sequences of events, or combinations of these.

When unstructured observations are the primary data source, the researcher faces the problem of imposing order on masses of data, usually in the form of field notes made during the period of observation. Consequently, attention must be given to editing one's notes, organizing them systematically, developing categories for classifying the data, and then counting and cross-tabulating the observations. Data collected under more controlled situations are usually more amenable to the analytical procedures used in quantitative social research generally.

REFERENCES

- BALES, ROBERT F., 1950. *Interaction Process Analysis*. Cambridge, Mass.: Addison-Wesley.
- 1970. *Personality and Impersonal Behavior*. New York: Holt, Rinehart, & Winston.
- BECKER, HOWARD S., 1970. "Problems of Inference and Proof in Participant Observation." Pp. 189–201 in William J. Filstead (ed.), *Qualitative Methodology*. Chicago: Markham.
- BUGENTAL, DAPHNE E., LEONORE R. LOVE, and ROBERT H. GIANETTO, 1971. "Perfidious Feminine Faces." *Journal of Personality and Social Psychology* 17 (March): 314–318.
- CARY, M. S., 1978. "The Role of Gaze in the Initiation of Conversation." *Social Psychology* 41 (September): 269–271.
- GOFFMAN, ERVING, 1963. *Behavior in Public Places*. Glencoe, Ill.: The Free Press.
- GOODE, WILLIAM J., and PAUL K. HATT, 1952. *Methods in Social Research*. New York: McGraw-Hill.
- HAMBLIN, ROBERT L., DAVID BUCKHOLDT, DONALD BUSHNELL, DESMOND ELLIS, and DANIEL FERRITOR, 1969. "Changing the Game from 'Get the Teacher' to 'Learn.'" *Trans-action* 6 (January): 20–36.
- HEYNS, ROGER W., and ALVIN F. ZANDER, 1966. "Observation of Group

- Behavior." Pp. 381-417 in Leon Festinger and Daniel Katz (eds.), *Research Methods in the Behavioral Sciences*. New York: Holt, Rinehart, & Winston.
- HUMPHREYS, LAUD, 1970. *Tearoom Trade: Impersonal Sex in Public Places*. Chicago: Aldine.
- JACOBS, H. A., 1967. "To Count a Crowd." *Columbia Journalism Review* 6: 37-40.
- KENEN, REGINA, 1982. "Soapsuds, Space, and Sociability: A Participant Observation of the Laundromat." *Urban Life* 11 (July): 163-183.
- LEWIS, OSCAR, 1959. *Five Families: Mexican Case Studies in the Culture of Poverty*. New York: Basic Books.
- 1961. *Children of Sanchez*. New York: Random House.
- MASSEY, M. E., 1970. "A Market Analysis of Sex-Oriented Materials in Denver, Colorado, August 1969—A Pilot Study." *Technical Reports of the Commission on Obscenity and Pornography*, Vol. 4. Washington, D.C.: U.S. Government Printing Office.
- NAWY, H., 1970. "The San Francisco Erotic Marketplace." *Technical Reports of the Commission on Obscenity and Pornography*, Vol. 4. Washington, D.C.: U.S. Government Printing Office.
- ROSENHAN, D. L., 1973. "On Being Sane in Insane Places." *Science* 179 (January 19): 250-258.
- SELLTIZ, CLAIRE, MARIE JAHODA, MORTON DEUTSCH, and STUART W. COOK, 1959. *Research Methods in Social Relations*. New York: Holt, Rinehart, & Winston.
- TEDESCHI, JAMES T., and SVENN LINDSKOLD, 1976. *Social Psychology: Interdependence, Interaction, and Influence*. New York: John Wiley.
- WHYTE, WILLIAM FOOTE, 1943. *Street Corner Society*. Chicago: University of Chicago Press.
- WINICK, C., 1970. "Some Observations of Patrons of Adult Theaters and Bookstores." *Technical Reports of the Commission on Obscenity and Pornography*, Vol. 4. Washington, D.C.: U.S. Government Printing Office.

Survey Research: Interview Studies

five

I. Introduction

II. Advantages of Survey Research

III. Disadvantages of Surveys

IV. The Interview

Types of interviews

Issues in designing an interview study

problems with purposeful misinformation

problems associated with the temporary role of respondent
the psychology of the respondent

problems from involuntary error

other sources of bias in interviewing

V. Creating an Interview Schedule

Communicating effectively

Common problems in question formulation

the double-barreled question

complex questions

the order of questions

probe questions

Using open or closed questions

VI. Pretesting the Interview Schedule

VII. Conducting the Interview

Selecting interviewers

Training interviewers

Approaching the respondent

VIII. Other Variations: The Telephone Interview

Advantages of telephone interviews

cost

sampling

convenience and access

supervision

Disadvantages of telephone interviews

length of interview and nature of materials to be covered

problems of interview termination

IX. Collecting Data by Interviewing: An Illustrative Example

X. Summary

I. INTRODUCTION

Survey research is the most widely used method of data collection in the social sciences. There are two main types of survey methods: the **interview** and the **questionnaire**. Each includes a number of subtypes. For example, interviews may be conducted in group settings, by telephone, or in a face-to-face private encounter between interviewer and respondent. The interview may be highly structured, with specific questions to be asked of all respondents, or it may be so unstructured that it resembles a conversation between friends more than an episode of data collection. Similarly, questionnaires may be administered to large groups in classrooms or other institutional settings, they may be mailed to respondents who fill them out in private and return them by mail, or they may be hand-delivered to respondents who are instructed to treat the questionnaire as if it were a self-administered interview.

While some social scientists may prefer other modes of data collection, virtually all social researchers are involved at some time in data collection by interviewing or questionnaire. Also much of the research reported in the media—the research findings offered to the typical “consumer” of research—will be based on survey data.

Given the pervasiveness of these procedures, we will devote two

entire chapters to data collection by survey. This chapter treats interviewing methods. The next will deal with questionnaires.

II. ADVANTAGES OF SURVEY RESEARCH

There are a number of important reasons why survey research methods, including interviewing and questionnaire studies, serve as a basis for a very large proportion of the research that is carried out by social scientists. For example, information on such topics as the attitudes and beliefs of large numbers of respondents is difficult to obtain by any other means. Political pollsters reflect this reality in periodic surveys of voters' attitudes about a variety of political issues and candidates; social scientists interested in religion do regular surveys about people's religious beliefs and practices; market researchers assess preferences of potential consumers; and radio announcers and disc jockeys do human-interest telephone interviews. While one can certainly make some assumptions about attitudes from the observation of behavior—i.e., from observing actual voting behavior, religious attendance, and purchasing habits—survey techniques are generally considered the most viable way of assessing attitudes on these topics.

Second, survey methods can be used to obtain information about events that have occurred previously and that now exist primarily in the memories of those to be studied. For example, respondents can be asked to provide information about their childhood experiences, about their medical histories, about past levels of participation in organizations, and so on. Unless these events have been recorded in some way—and even when audio or visual recordings, photographs, or archival records such as letters or journal entries are available—the collection of *retrospective* data by survey provides detail, interpretation, and information about situational or contextual factors unavailable via any other research technique.

A third advantage of survey methods is that they permit the collection of data from large numbers of respondents in relatively short periods and at relatively low costs. National and regional surveys are regularly conducted in this and other countries to provide information on such things as income and educational attainment, changes in political preferences, childrearing practices, sexual behavior, and a variety of other topics of interest to politicians, legislators, administrators, social scientists, and the general public. The amount of information obtained in surveys is broader than what could be obtained through any other procedure. For example, a national survey of members of the Lutheran Church used a questionnaire that required over three hours to complete. Using this instrument, the researchers obtained extensive and detailed information from a represent-

ative sample of approximately 6,000 Lutherans residing throughout the United States (Strommen et al., 1972).

Because data can be collected from large samples of respondents at comparatively low costs, surveys are frequently used when the researchers desire to generalize their findings from a sample to a larger population. Whereas it is most difficult to conduct experiments or to do participant observation with anything but a comparatively small group of respondents, survey research can be conducted with large samples and the consequent ability of the researcher to generalize from the data to the larger population is increased.

Finally, because of pervasiveness of surveys, it is generally relatively easy to get people to participate. Political pollsters typically obtain very high response rates and, although response rates decrease as the survey becomes more complex, personal, or lengthy, the survey remains the most effective procedure for obtaining data from large samples.

III. DISADVANTAGES OF SURVEYS

The fact that interviews and questionnaires provide the best (and in some instances perhaps the only) method for collecting certain types of data does not mean that these are the best research tools *period*. We have continually emphasized that the researcher must choose the method or methods that are called for by his or her particular research problem and must make this decision within the constraints imposed by the resources available.

A primary disadvantage of the survey is that it is virtually impossible to test for cause-effect relationships with survey methods. As a tool for causal analysis or for careful testing of theory, then, the survey is weak when compared with the experiment. In addition, the depth of information that one is able to obtain is usually not nearly so great as is the case with participant observation. Nor is the researcher likely to learn as much about important current situational or contextual factors that affect the behavior of the respondent. What this all means is that while survey research has important advantages, it also has many disadvantages.

IV. THE INTERVIEW

While the student who is just beginning his or her training in research methods may have had relatively limited experience with the research interview, most young adults in our society have had considerable experience with interviews of some kind. For example, when we apply for a job, we are often interviewed by prospective employers about our personal backgrounds and experiences. The process of admission to college often

includes an interview with a university official. Nurses and physicians collect medical histories from their patients, and ministers or counselors may interview those who come to them for help or advice. The interview as applied in social science research may differ in format or objective from some of these other kinds of interviews, but many of the principles of conducting a successful interview are the same whether the context is a research project or a lawyer-client exchange.

Cannell and Kahn (1968: 527-528) define the research interview as "a two-person conversation, initiated by the interviewer for the specific purpose of obtaining research-relevant information, and focused by him on content specified by research objectives of systematic description, prediction, or explanation." These authors state that while interviewing to obtain information is used by many public and private organizations, its most ambitious and demanding use is usually found in the social sciences, where lengthy interviews on subjects that raise difficult problems of recall, potential embarrassment, or self-awareness are not uncommon. Among the classic studies that have used interviewing as the primary mode of data collection are explorations of consumer behavior, fertility and family planning, sexual behavior, mental health and illness, political behavior, supervisor-subordinate relations, and attitudes of workers.

Interviewing, then, is one of the most basic forms of data gathering. In some circles, it has virtually been treated as synonymous with sociological methodology:

Sociology has become the science of the interview. . . . The several branches of social study are distinguished from one another perhaps more by their predilection for certain kinds of data and certain instruments for digging them up than by their logic. . . . Sociologists have become mainly students of living people. Some, to be sure, do still study documents. Some observe people *in-situ*; others experiment on them and look at them literally *in vitro*. But, by and large, the sociologist of North America, and in a slightly less degree in other countries, has become an interviewer. The interview is his tool; his works bear the marks of it. (Benney and Hughes, 1956: 137)

However, as is true with all of the other techniques that we will discuss, interviewing is a phase of the research process and not the entire process itself. Cannell and Kahn describe the total process as a series of discrete steps:

1. Creating or selecting an interview schedule (set of questions, statements, pictures, or other stimuli to evoke response) and a set of rules or procedures for using the schedule.
2. Conducting the interview (that is, evoking the responses or events that are to be classified).
3. Recording these responses (by means of paper-and-pencil notes, electronic equipment, or other devices).

4. Creating a numerical code (that is, a scale or other systems of numbers into which the recorded responses are to be translated, and a set of rules for making the translation).
5. Coding the interview responses. (1958: 531)

These five steps are then followed by data analysis and report preparation. We have discussed the tasks of creating codes, coding data, analysis, and report preparation in other chapters. Here our focus is primarily on the tasks of developing questions, conducting interviews, and recording responses.

Types of Interviews

One of the major differences that exists among interviews is the amount of structure that the researcher imposes on the respondent. Interviews range along a continuum from the highly structured interview schedule which permits no deviation to the largely unstructured, undirected exploratory interviews. Highly **structured interviews** will usually contain a series of very specific questions that are to be read to the respondent, along with a set of predetermined response categories. The working form and order of the questions is designed to be exactly the same for all respondents. The respondent simply selects one of the answers provided and the interviewer records that response in the appropriate place on the interview schedule. Few, if any, **open-ended** items or questions are used.

At the other end of the continuum is the **exploratory interview**, in which the interviewer is to explore a variety of preselected topics with the respondents but with little concern for asking specific questions in any preestablished format or sequence. Using this approach, the interviewer does not have a standard set of questions that are to be asked of all respondents nor is any attention given to developing response categories for the subject. Rather, "the interviewer explores many facets of his interviewee's concerns, treating subjects as they come up in conversation, pursuing interesting leads, allowing his imagination and ingenuity full rein as he tries to develop new hypotheses and test them in the course of the interview" (Becker and Geer, 1957: 28).

Between these two extremes are a variety of other ways of conducting interview studies. Near the more structured pole of the continuum might be an interview that includes specific questions but asks them in a largely open-ended format. That is, questions but not response categories are predetermined. The respondents are all asked the same questions but are given their freedom in answering them in the manner they choose. The researcher then is faced with the responsibility of coding the responses into categories for analysis. Nearer the other pole is the situation where the researcher has some rather specific topics that are to be covered, and

these are included in an **interview guide**. However, the exact manner in which the questions are asked and their sequence are determined in the course of the interview itself. The guide is used to make sure that all of the issues of concern receive attention during the course of the encounter but the interview itself remains unstructured.

The type of instrument one chooses will be determined by the particular research needs and the purposes of the research. Maccoby and Maccoby (1954), along with several others, have suggested that the highly structured interview format is best suited for more specific hypothesis testing and the rigorous quantification of results. This format also assumes that the researcher has rather extensive information already available about the subject and about the respondents to be interviewed such that meaningful questions and response categories can be formulated before the researcher goes into the field (though we assume some extensive pretesting will have been necessary in developing both the questions and the response categories). A more highly structured format would also be selected if the researcher desired to obtain the same basic set of information from all subjects and if a large sample of respondents was to be included in the study. In terms of the latter, a structured format certainly facilitates data handling and analysis. The unstructured format, on the other hand, is best suited for exploratory studies and for studies in which detailed information might be needed on more complex and detailed issues.

Issues in Designing an Interview Study

Despite its importance and widespread use in the social sciences, the collection of data by interviews carries some serious liabilities. As Nelson, Bechtol, and Johnson (1977: 94) have summarized, the results one obtains from an interview study might be anything from highly reliable, scientifically valid and valuable data to virtually worthless information. Whether one's study falls toward the positive or the negative pole of the continuum will depend upon a number of important factors.

A number of years ago Vidich and Bensman (1954) identified several important types of errors and sources of misinformation in interviewing. Their list included: (1) errors resulting from purposeful intent on the part of the respondent to deceive or mislead, (2) problems associated with the temporary role of the respondent, (3) errors related to the psychological state of the respondent, and (4) involuntary error. We will talk about each of these in some detail.

Problems with purposeful misinformation In their original article, Vidich and Bensman suggested that purposeful misinformation may result from any of the following factors:

1. Biased or slanted information resulting from a respondent's attempt to influence the results of the research. Such biases might be expected among community leaders who are concerned with giving a favorable impression of their town.
2. Dramatized information designed to make the informant and the community seem less prosaic.
3. Over-information given by reformers who want to use the research to expose and reform the community.
4. Blockages of attempts to gain information about the dynamics of certain institutional complexes such as sex, power, and class. . . .
5. Rationalizations of publicly unacceptable behavior. . . .
6. Information distorted to serve personal ambitions, self-aggrandizement, self-protection or to serve in working out personal feuds.
7. Advance preparation of responses based on rumors and other types of intercommunication about the research, leading to stylized and stereotyped responses. (1954: 21)

Any or all of these pressures toward conscious inaccuracy by respondents must be seen as possible sources of error in interview research. An interview is a form of social exchange between two or more persons. The respondent will typically be motivated to project a certain image of self, family, social group, community, or region. That image, at least in part, will be a function of the quality of the interaction between the interviewer and respondent and will have a direct impact on whether respondents are candid in their answers to questions they define as "sensitive."

There is evidence that conscious bias is most likely in interviews where there is considerable social distance between interviewer and respondent (Williams, 1968). That is, lower-class respondents or members of disadvantaged minorities may be less willing to be truthful when they are interviewed by those who are from upper-status or majority populations. The greater the perceived social class difference between respondent and interviewer, the less likely it is that the interviewer will be given detailed, accurate answers to sensitive or threatening questions (Schuman and Converse, 1971; Campbell, 1981).

There are no easy solutions to this problem. Most handbooks on interviewing say that the interviewers must establish **rapport** with their subjects if they are to be effective. High rapport with the respondent is seen as the most effective way to dispel a subject's fear and to put him or her at ease so that questions will be answered completely and honestly. At the same time, too much rapport may be a bad thing. If the respondent views the interviewer as friendly and attractive, he or she may try to answer the questions in ways that the respondent thinks will please the interviewer or at least will be consistent with the interviewer's values and attitudes as perceived by the respondent (Williams, 1968). Well-trained interviewers try to minimize this problem by not giving respondents any cues about

their own attitudes on the topic under study. Such impartiality or neutrality does not eliminate the possibility that subjects will try to guess the attitudes of the interviewer, but the biases associated with such guessing are more likely to be random, or to cancel each other out, than if interviewers reveal their own biases.

The ability of interviewers to obtain accurate information about topics that are embarrassing or threatening to the respondent is also problematic. However, there is some evidence that when properly used—when items are worded appropriately and interviewers conduct the interview as specified with respect to both technical competence and proper emotional/interpersonal climate—interviews can be used to provide reliable and valid data on topics as sensitive as personal finances, sexual behavior, drug use, and other criminal or deviant behavior (cf. Bale, 1979).

John Ball (1967) conducted an extensive test of the reliability of information obtained during interviews from narcotic drug addicts. Many critics had said that deviant groups, particularly persons who had been engaged in illegal behavior, would be unwilling to report their proscribed behavior honestly. Ball's sample included 59 Puerto Rican addicts who had formerly been incarcerated at the U.S. Public Health Service Hospital in Lexington, Kentucky. The specific procedure he used to assess the reliability and validity of the interview data from the addicts was to compare specific interview items with three other data sources: (1) clinical and administrative records on the individual which were kept at the hospital, (2) FBI arrest records, and (3) urine samples that were obtained from the subjects at the time of the interview. The interview items that could be checked against these other sources of information included age, age at onset of drug use, type and place of first arrest, total number of arrests, and presence or absence of drug use at the time of the interview.

These comparisons increased the confidence of the researchers in the validity of interview data from deviant populations. In 82.3 percent of the cases the reported age of the subject given at the time of the interview (often many years after the hospitalization experience) was the same as the age recorded on the hospital records. In all but two instances where there were discrepancies, the discrepancies were of only one or two years. Age at onset of drug use was consistent between the hospital record and the interview in 65.3 percent of the cases and was within one to three years in another 27.3 percent of the cases. In 80.7 percent of the cases the interview report of first arrest was consistent with the subject's FBI record. Some subjects had forgotten the details of their arrest history but most were willing to answer the "sensitive" items, and their answers were generally consistent with data on the same topics obtained from institutional records or other sources.

Over 70 percent of the subjects gave valid reports of their criminal history, and those who did not tended to underreport *minor* rather than

major offenses. As for the validity of current reports of drug use, Ball reported:

Of these 25 subjects "on the street" in Puerto Rico at the time of interview and urinalysis, eighteen reported that they were not using heroin and their urinalysis was negative for opiates; five admitted drug use and their specimen was positive; and two denied drug use, but the laboratory report was positive. On this basis, it may be said that 92 percent of the subjects' reports of current drug use were valid, employing the criterion of chemical analysis. (Ball, 1967: 652-653)

Ball concluded that, considering the many sources of response error, the interview data could be considered quite reliable and valid.

Research by one of the present authors (Bahr and Houts, 1971) into the reliability of responses to interviewers by skid row men produced similar evidence that interviews could be used to obtain reasonably accurate information from deviant populations. Although the stereotype of the skid row man does not include the trait of honesty, careful checks for discrepancies between responses given to interviewers on New York City's Bowery and client records at municipal shelter and welfare offices revealed that the responses of homeless men who agreed to be interviewed were no more likely to be inconsistent with available agency records than were responses from other low-income respondents. However, because homeless men tend to be aged and to some degree personally disoriented, with high rates of both physical and mental illness, their ability to give accurate answers was shown to be directly related to the complexity of the information sought and the clarity of the items in the interview schedule. There was also evidence that the skid row men recalled recent events more accurately than past events, and that they, like everyone else, were influenced by social desirability considerations: that is, given an ambiguous question, they were likely to answer in a manner that put themselves in a good rather than a bad light. However, when the questions were unambiguous, there was no evidence that the social desirability biases in the responses of skid row men were any more frequent than in other populations. The article entitled "Can You Trust a Homeless Man?" concluded with the positive finding that interview data from skid row residents was about as reliable as interview data from other people.

Problems associated with the temporary role of respondent In most interview situations the interviewer and the respondent do not know each other prior to the interview, and it is probable that they will never see each other again after the interview. As Denzin (1970: 135) has noted, "There is nothing intrinsic in this fleeting relationship that will force the respondent to behave as he is supposed to." Given the temporary nature of the exchange, the respondent may have little motivation either to participate

in the first place or to give complete and accurate answers. The interviewer must somehow motivate the respondent to participate and give appropriate answers.

Some studies have provided financial incentives to get people to participate in an interview, but most researchers appeal to other motivations. Kahn and Cannell (1957) suggest that respondents participate for two reasons: intrinsic motivation that results from the fact that the experience and the relationship with the interviewer is valued by the respondent, and "instrumental" motivation that grows out of the fact that the respondent sees the effort of which the interview is a part as congruent with personal values and goals. Kinsey (1948) emphasized altruistic motivation as a reason for respondents to participate in his detailed, very personal studies of human sexuality. Strommen et al. (1972), in their study of the religious behavior of Lutherans, emphasized in their introduction to their respondents that "We hope it makes you feel as good as it makes us feel to be helping with the biggest and most comprehensive study of the attitudes and life styles of Lutherans ever made—a study of great significance, whose value will continue for many years to come."

Whatever the type of motivation, the important point is that the interview is a fleeting relationship, and unless the respondent can in some way be convinced of the importance of the study and of his or her own role as an honest informant in that study, incomplete, inaccurate, and biased data will likely be the result.

The psychology of the respondent Interview respondents, like everyone else, are subject to a variety of personal characteristics and attributes that affect the accuracy and quality of their responses (Vidich and Bensman, 1954: 24). These characteristics include things like **attitudinal set**, individual and collective illusions and myths about their particular setting and time, unusual motivations of interest and disinterest, personal fears and anxieties, and so on.

In their experimental work on improving the accuracy of interviewee response, Cannell, Oksenberg, and Converse (1977) identified several psychological or "personal make-up" factors that may explain poor performance by a respondent. A critical factor is that many studies do not adequately define for the respondent exactly what his or her role should be. For example, the researchers demonstrated that poor performance tended to show up when the respondents did not know what was expected of them or when they did not understand the goals of the study or how they were supposed to deal with particular questions. In such cases, respondents tend to tolerate an interview but, lacking a clear understanding of their task, may produce unreliable information.

Cannell and his colleagues (1977: 309) report that "the evidence suggests that most respondents accurately report information that is readily

accessible and nonthreatening. As tasks become more demanding, however, respondents do not work hard enough to retrieve information from memory and to organize information efficiently, and they do not accept even a minimal risk of embarrassment."

To overcome some of these problems, it is recommended that survey research should:

1. Clarify for the respondent what is expected of him or her.
2. Provide cues as to how to be most efficient in answering particular questions.
3. Motivate the respondent to work diligently to recall and organize information and to report even potentially embarrassing material.

Three characteristics are required of a "good" interview subject, according to Cannell et al. First, the subject must be able to comprehend the question being asked and to recall and process the information necessary to give an answer. Second, the subject must be willing to exert the necessary effort to do this. Third, because accurate and complete answers often require frank disclosure, the subject must be willing to provide that information, that is, to trust the interviewer.

In addition to these aspects of motivation, the researcher must constantly be alert for possible biases associated with such things as attitudinal set ("yea-saying") and fears and anxieties that might be generated in the respondent by the questions asked. Both of these issues are treated in greater detail later when we discuss the construction of interview schedules and questionnaires.

Problems from involuntary error Respondents are frequently unable to provide the information that is wanted, not because they are unwilling to do so or because they wish to deceive, but simply because of lack of information, disorientation, memory loss, or fatigue (Vidich and Bensman, 1954; Bahr and Houts, 1971). Hanson (1979) notes that many respondents make mistakes even when providing the most elementary facts (such as the ages of their children or their own year of birth). An even more important problem, however, emerges when responses are used as indicators of subjective states such as feelings, attitudes, or beliefs. Hanson argues that indicators of subjective states commonly used in survey research are, at best, crude. Subjective states are highly changeable and are often difficult to report even by the most conscientious respondent. Furthermore, because subjective states are highly changeable and situation-specific, responses to indicators of such states are not likely to be good predictors of other events or behaviors even when the respondent has conscientiously tried to answer the items fully and honestly.

There are many other reasons why a respondent might give incom-

plete or inaccurate information in an interview without intending to deceive or mislead. For example, when factual information is sought, the person may not know the facts being talked about. Some wives may not know how much money their husbands make or even many details about their husbands' employment. Respondents may not be familiar with a particular political issue or policy decision, although the interview schedule assumes such familiarity. In such cases, answers given are likely to be incomplete, inaccurate, or both. Similarly, questions about "typical" events or experiences may not be meaningful to some people. For such issues, a properly constructed interview schedule would include questions that would screen out persons to whom certain questions did not apply. In interviews dealing with such issues as parent-child relationships, respondents who have not had children, or whose children are grown and have left the parental home, may find the items inappropriate or meaningless but they may still, as "good respondents," answer.

Attitudinal items are based on the assumption that everyone has attitudes about almost any topic. Many respondents may never have thought much about, for example, a proposed development of a new nuclear weapons system. However, rather than appear ignorant, they will readily give an answer when asked. Such "top of the head" responses are clearly less significant than those responses from people who have devoted much thought to the topic.

Even if the respondent has the necessary information, possesses the experience, and has developed an identifiable attitude about some issue, two other problems may impede the collection of given data, namely the problems of recall and of distortion or repression. First, the problem of recall: the event may have been forgotten, or the details of an experience may have faded from memory. Cannell, Oksenberg, and Converse (1977: 306-307) suggest that the mental material accessible to a respondent who is motivated to answer depends on a complex interplay of factors, including:

1. *Time lapse.* As the time between an event and the interview increases, there is increased underreporting of information about the event. This finding is not an unexpected one, but the rapidity of memory loss is surprising. For example, failures to report to an interviewer a visit that the respondent had made to a physician doubled over a two-week period, from 15 percent in interviews held one week after the visit to 30 percent in those held two weeks after the visit.
2. *Salience.* Events important to the respondent are reported more completely and accurately than others.
3. *Social desirability.* The reporting of an event is likely to be distorted in a socially desirable direction. For example, events or characteristics defined as embarrassing, sensitive, threatening, or incongruent with one's public image may not be reported, or if reported, may be distorted in ways that make the

respondent "look good." Such "social desirability" biases may occur without any conscious effort by the respondent to be deceptive.

4. *Event-specific recall.* There is more variation in the way different events are recalled by a given observer than in the recollection of a single event by multiple observers. That is, although some respondents may be more accurate than others, generally, most respondents report some information accurately and some poorly. (Cannell, Oksenberg, and Converse, 1977: 307)

A related problem has to do with distortion or repression of a previous event or experience. Some distortion is conscious, designed to enhance the image of the respondent in the interviewer's eyes, but descriptions of other events, particularly if they were painful or traumatic, may be incomplete or distorted because the respondent has repressed or unconsciously reinterpreted the happenings being studied.

Other sources of bias in interviewing Among the other sources of bias in data collection by interview are the personal characteristics of the interviewer and how effectively he or she performs the role. Some early studies showed a strong relationship between the attitudes of the interviewer and the responses obtained from the people interviewed (Rice, 1929). To some degree, interviewers tend to slant people's answers in the direction of the interviewer's opinions (Cannell and Kahn, 1968). Most interviewer training programs are designed to decrease the impact of interviewer attitudes on respondents. However, despite careful training, interviewer attitudes continue to be a source of bias. In a recent study by one of the authors, three Native American interviewers were hired to interview a sample of tribal members on their attitudes toward proposed energy development. The interviewers were thoroughly trained, and they conducted practice interviews that were evaluated for bias and used as aids in teaching interviewer objectivity and neutrality. Despite this effort, when the data were analyzed we found that the one best predictor of respondents' attitudes about energy development was the attitude of the interviewer. Most respondents fell in one of three categories, and placement in each category was largely determined by which of the three staff members had conducted the interview.

In addition to personal attitudes, other interviewer characteristics are possible sources of bias. For instance, the visible ethnic or gender characteristics of the interviewer may suggest to a potential respondent stereotypes that influence answers to questions (Cannell and Kahn, 1968: 549). Such imputed background characteristics as age, gender, religion, race, or education may trigger certain attitudes or predispositions among respondents. Respondents who want to make a good impression on the interviewer might answer a question on sex discrimination in employment quite differently when the question is asked by a woman. Similarly, white interviewers may get different answers from black respondents on questions

about racial discrimination and prejudice than would black interviewers (Schuman and Converse, 1971).

Cannell and Kahn summarize:

... background factors are important in the interview because they constitute a kind of subsoil in which many of an individual's attitudes, motives, and perceptions have direct roots. But the background characteristics of each participant in the interview have additional importance because they provide cues for the other participant. Certain attitudes, motives, and stereotypes are triggered in the respondent's mind by his perception that the interviewer possesses certain background characteristics. The interviewer may be influenced in the same fashion by his initial perceptions of the respondent. Such reactions may in turn influence the behavior of both participants. (Cannell and Kahn, 1968: 550)

The various sources of bias in interview research can never be completely eliminated, but there are some procedures that decrease the probability of serious bias. For example, the most essential procedure to reduce potential bias associated with the personal attitudes of interviewers is a carefully designed procedure for selecting interviewers, combined with a comprehensive training program. Most survey research organizations have written handbooks and hold training workshops to teach interviewers how to minimize the effects of their personal characteristics on the data they collect.

Another way to reduce bias in interviewing is to frame the questions clearly and unambiguously. People do not all use language in the same way, and the meaning of words and phrases varies from one context to another. As we concluded in an earlier study of bias and discrepancies in interview data, the "number of invalid responses can be reduced far more efficiently by greater scrutiny of the kind of question being asked than by lengthy programs aimed at producing 'high rapport' interviews" (Bahr and Houts, 1971: 382). The wording and placement of questions in the interview schedule determine the quality of the responses, and we discuss it in greater detail later in this chapter.

In addition to asking the right questions in the right way, and properly training and supervising interviewers, there are some other things that can be done to reduce error in interview data. It is essential that the researcher know the subjects and their social context and be familiar with the interplay between that context and the topics being studied. An interviewer who knows the people and the topic and knows how they are typically related can often identify exaggerations or misrepresentations. Sometimes it is appropriate to ask the respondent about apparent misrepresentations. Other times a good interviewer will note the apparent inconsistencies so that they can be taken into account during the coding and analysis of the data.

Problems of misinformation can sometimes be detected by asking the

same question in a slightly different form later in the interview. Interview data can also be supplemented and occasionally verified by comparisons with data from other sources, such as agency records or newspaper accounts. If a respondent understands that others are to be interviewed on the same topic, or that the researcher is also checking other data sources, he or she may be less likely to try to mislead or to give incorrect answers knowingly.

Sometimes interviewers inspire confidence and accuracy of response by clearly establishing the importance of the project and telling the respondent how critical honest answers are to its successful completion. Earlier, in the discussion of motivating people to participate in an interview, reference was made to the fact that in their presentation of self, interviewers can create the impression that what they are about is important and is worth the respondent's time and interest. Interviewers can further increase a potential respondent's confidence in them by careful selection of role, dress, and language. In this regard, Denzin (1970: 140) advised interviewers to: "Dress in the mode of dress most acceptable to those being interviewed but employ a style that communicates who *you* are with respect to *them*." That is, the temporary nature of the encounter between the interviewer and the respondent need not prevent rapport and honest communication if the interviewer is properly prepared and conducts the interview in a professional way.

Uniformity in interviewer performance can be increased by means of a highly structured interview schedule. The greater the number of closed-format questions, the lower interviewer improvisation ought to be. Specific instructions on the interview schedule may also serve to prompt interviewers to probe for additional information in a consistent way and can even systematize the verbal reinforcement interviewers transmit to keep the respondents motivated.

Another way to improve the quality of the interview is to select interviewers in a way that undesirable interaction with respondents is minimized (Cannell and Kahn, 1968). For example, if characteristics of the interviewer such as age, sex, or race appear relevant, then the interviewers hired should be those whose personal characteristics minimize the probability of conflict or problems in building rapport. For example, young women might be selected as interviewers of women residing in urban centers because they are less likely than young men to be seen as threatening to a potential respondent.

Cannell and Kahn (1968: 562) have proposed a series of techniques to help researchers overcome problems of involuntary error associated with memory lapses and the like. The recommended techniques include: (1) asking the respondent to consult records that may be available about the event or experience (such records might include bank deposits, income tax forms, diaries, and so on); (2) providing the respondent a context that helps

him or her to reconstruct the past, such as locating events in real historical and geographic time and space; (3) simply alerting the respondent to the problem of bias in recall and urging that great care be taken in answering the questions; (4) wording questions so that they require recognition rather than recall: the respondents might be given a list of possible answers and then asked to select the one from the list that best fits his or her situation.

V. CREATING AN INTERVIEW SCHEDULE

Following Cannell and Kahn (1968), we have defined the research interview as a type of conversation that is initiated explicitly for the purpose of obtaining information. To get information, we usually must ask questions. Consequently, developing the proper questions is a critical part of any survey research project. Questions must be worded so that they will provide the necessary data, and they must be asked in ways that motivate respondents to answer fully and honestly. If the wrong question is asked, or if a question is asked in such a way that the subject either cannot give an appropriate answer or is not motivated to do so, then the interview will not provide reliable data.

Denzin provides the following guidelines for formulating questions:

... questions should accurately convey meaning to the respondent; they should motivate him to become involved and to communicate clearly his attitudes and opinions; they should be clear enough so that the interviewer can easily convey meaning to the respondent; they should be precise enough to exactly convey what is expected of the respondent . . . ; any specific question should have as a goal the discerning of a response pattern that clearly fits the broad intents of the investigation . . . ; if questions raise the possibility of the respondent's lying or fabricating (which is always a possibility), care should be taken to include questions that catch him up, or reveal to him and the interviewer that his previous answers have been incorrect. (Denzin, 1970: 129)

Communicating Effectively

The central problem in asking questions is that of adequate communication. The researcher must clearly communicate to the respondent what he or she wants to know. The interviewer's language must be understandable to the respondent; at best, the interview will be conducted at the level of language most comfortable to the respondent:

Communication can occur across the boundaries of age and background, but the form and content of questions is limited by the shared vocabulary between researcher and respondent. How large that shared vocabulary must be will depend on the subject matter, complexity, and conceptual level of the topics under investigation. The issue is not the size of the shared vocabulary, but its adequacy for communicating material required by the research objectives. (Cannell and Kahn, 1968: 554)

Becker and Geer (1957: 28–29) have also noted the importance of knowing the language of the interviewee *both* in order to communicate what it is that one is interested in knowing and in correctly interpreting what the respondent reports back. They note:

(A)lthough we speak one language and share in many ways in one culture, we cannot assume that we understand precisely what another person, speaking as a member of such a group, means by any particular word. In interviewing members of groups other than our own, then, we are in somewhat the same position as the anthropologist who must learn a primitive language, with the important difference that, as Icheiser has put it, we often do not understand that we do not understand and are thus likely to make errors in interpreting what is said to us. (Becker and Geer, 1957: 28)

To minimize potential problems of communication in studies of the general public or of disadvantaged populations, many analysts propose that the words and ideas in an interview schedule must be simplified to the level of the least sophisticated of potential respondents. However, such simplification may create as well as solve problems: The more sophisticated respondents may react negatively to questions asked in very simple language. This problem is less critical when one's respondents belong to a homogeneous subculture, as, for example, when one interviews a group of medical students. However, when interviewing a cross-section of subjects on the same topic, one must take into account the various levels of sophistication in formulating questions that represent some "common denominator" of comprehensibility.

Perhaps the most important point is that the researcher must take the necessary time to understand the backgrounds and abilities of the potential respondents. With adequate preparation, researchers will have some familiarity with local definitions and practices, abilities and interest. Questions can be too difficult, but they can also be overly simplistic, and either extreme is likely to offend some segment of one's sample. The best way to estimate the appropriate level of difficulty and to improve one's chances for building rapport rather than losing it during the interview is to do one's homework carefully before going into the field and then to pretest one's instrument and, if possible, do a small-scale pilot study before proceeding with the final full-scale field work (see Schuman and Presser, 1977; 1981).

Common Problems in Question Formulation

Even if one knows "the territory" sufficiently to ask meaningful questions, there are still some common mistakes that should be avoided. Many of these we consider at greater length in the next chapter on questionnaires, but let us introduce them here because they apply to both interview and questionnaire studies.

The double-barreled question The double-barreled question is a common problem in interviews or questionnaire research. This type of question asks a respondent to respond to two issues (often quite unrelated) at the same time. For example, the question "Would you like to see the United States expand its nuclear missile capabilities in Western Europe or concentrate on the development of an anti-terrorist strike force?" asks about expansion or concentration (one issue) and about nuclear versus anti-terrorist capabilities (another issue). One might favor both the expansion of nuclear missile weaponry in Europe and the development of anti-terrorist capability. One could also oppose both. The only appropriate solution to this impossible query is to break it into two separate questions. People will respond to double-barreled questions; the problem is that the analyst cannot interpret what a response means.

Complex questions Particularly in studies of the general population or of disadvantaged populations it is wise to avoid long, complex questions. A respondent usually has only a moment to consider how to answer, and in that limited time it is difficult to express one's opinion about a complex issue. On the other hand, if one uses an unstructured, open question format where there is little concern with uniformity across respondents, then difficult issues can be covered because an interviewer has the time to listen and usually is instructed to continue questioning until a topic is fully explored.

When complex questions are necessary, it is usually wise to give respondents copies of the response alternatives. For example, a respondent may be given a card with the response alternatives at the time the interviewer asks the question. Then the respondent chooses from the list of structured responses and is aided in answering by the visual list; there is no need to try to remember everything the interviewer said. Response cards are also useful in asking sensitive questions. If the respondent is asked to report annual income, a card from which a category can be selected is less of an invasion of privacy than would be asking for the specific amount of personal income. The income category might be identified by a letter or a number so that the respondent can use the letter or number in answering the question and not even have to spell out a specific category in dollar terms.

The order of questions Research on the results of arranging or ordering questions has revealed that the interview should begin with questions that are interesting to the respondent and at the same time are non-threatening and relatively easy to answer. Once rapport and interest have been established and the pattern of question and answer has become "natural," the interviewer can proceed to the more complex or sensitive issues. Thus, one might begin with a brief introductory statement describing what

the study is all about, why and how the respondent was selected, the contribution the study will make, and why the respondent's participation is critical. The initial questions might deal with such straightforward matters as length of residence in the community, size of family and ages of children, and so on. Later questions would focus on matters more central to the project's objective. For example, an interview schedule on religion might reserve the more personal questions about religious beliefs and personal spiritual experiences for the later phases of the interview.

Probe questions It is almost always necessary to buttress central questions with supplementary **probes** which guarantee, to some extent, that even taciturn respondents will provide the essential minimum of detail. For example, following an initial response, the interviewer might ask "Is there anything else that you can tell me about that?" or "Why do you feel that this is the case?" or "Are there any other reasons that you can identify?" Probe questions often yield the underlying essentials that a poorly constructed interview schedule might miss entirely. They encourage the respondent to think more deeply about an issue or to expand or explain a preliminary response. A well-trained interviewer will know when to ask a probe question even if one is not called for on the instrument. In large-scale studies the probes should be asked uniformly and at specified points so that all respondents have essentially the same "stimulus" presented for response. If only certain respondents receive probe questions, serious problems of comparability and missing data are created. In more exploratory work interviewers are freer to probe as they feel it is appropriate. However, in standard instruments for large samples it is usually wise to specify clearly in the interview schedule when and where probe questions are to be asked.

Using Open or Closed Questions

An aspect of interview structure is the question of open versus closed categories of response. In the literature on survey research much attention has been given to the advantages and disadvantages of each type of response category. With the **open-ended** option, respondents are encouraged to answer in their own words and to reveal their own definitions of the situation. The interviewer's responsibility is to ask the question and to probe until the respondent has finished the relevant detail, and to record that detail as carefully and fully as possible.

Closed response categories require the respondent to select from among a series of alternative answers provided by the researcher. Under the closed-response format, the researcher maintains control over the form, length, and content of the possible answer (Cannell and Kahn, 1968). The fixed-response or closed-category option minimizes the problems of coding. It

does so, however, at the risk of imposing the researcher's view of reality upon the respondent and forcing the respondent to reply within the limits previously established by the researcher, whether they fit or not (see Schuman and Presser, 1979).

These problems and advantages can be illustrated by alternative forms of a question on why people had decided to stop coming to church:

Open-ended format: Would you please tell me why you do not attend religious services?

Close-ended format: Which of the statements below best describes your reasons for not attending religious services? (check one)

1. When I grew up and started making decisions on my own, I stopped attending church.
2. I found other interests and activities that led me to spend less time in church-related activities.
3. The church no longer was a help to me in finding the meaning and purpose of my life.
4. No particular reason; I just got away from the church and never got involved again.

Data from the closed format questions are much easier to analyze. However, with the closed-category option, one is never sure that the set of responses from which the respondent chose actually contained the reason the respondent would have given had the question been asked in the open format mode.

According to Cannell and Kahn (1968: 565), there are five considerations relevant to choosing between open-ended and closed questions: (1) interview objectives, (2) respondent information level, (3) structure of respondent opinions, (4) respondent motivation to communicate, and (5) initial interviewer knowledge of the preceding respondent characteristics.

Typically, the more knowledge one has about an issue, the safer it is to ask questions in a closed-response format. The less one knows, the greater the risk of error in using a closed-response format. Frequently, one can move from an open- to a closed-response format during a well-developed pretest and pilot study. In the pretest the researcher might leave most of the questions open-ended; then, on the basis of the types of responses obtained, a preliminary structured instrument can be created and fielded with a small sample. The pilot study should involve enough respondents to allow the researcher to anticipate the range and distribution of responses in the major survey, and the closed-category responses should be designed accordingly. Whatever the approach one chooses, however, a careful pretest is always necessary, and frequently a pilot study is an essential stepping-stone to the large-scale sample survey (Schuman and Presser, 1979).

VI. PRETESTING THE INTERVIEW SCHEDULE

Whether one uses an open or a closed format, whether one's interview schedule is structured or unstructured, and whether the interview is conducted person-to-person or by telephone, it is always important to **pretest** the instrument before beginning the study. A pretest is a vital part of virtually any type of research. Basically, the pretest involves testing one's instrument and procedures on a small scale and then redesigning them to correct errors or problems revealed. Sometimes several pretests are needed before an instrument is suitable for use in large-scale research.

The pretest for an interview study should answer several questions:

1. Has the researcher included all of the questions necessary to test the research hypothesis? Even the most careful researcher will sometimes find that an essential variable has been overlooked. A well-designed pretest followed by a preliminary analysis of the apparent effectiveness of the collection procedures and of the kinds of data obtained will usually answer this question.
2. Do the questions asked elicit the types of response that were anticipated? Is the researcher's intent adequately communicated to the respondent? If not, some adjustments must be made and perhaps new questions added or substituted.
3. Is the language of the research instrument meaningful to the respondents? Among the relevant questions here are: Is the level of the vocabulary appropriate for the respondents? Do they understand the questions? Are the questions worded in ways that are consistent with local or subcultural usage? Can the questions be understood and answered without further explanation or rewording?
4. Are there other problems with the questions, such as double meaning or multiple issues imbedded in a single question?
5. Finally, does the interview guide, as developed, help to motivate respondents to participate in the study?

A pretest is really a project in miniature. Responses should be coded and submitted to analysis. If this is done, problems and pitfalls can be identified and eliminated before major resources have been committed to procedures that prove to be inefficient or misleading.

VII. CONDUCTING THE INTERVIEW

The task of actually collecting interview data depends, in the end, on the interviewer. The interviewer's work is much simpler if the project leaders have done their work well. That is, the instrument the interviewer finally uses should contain carefully framed questions that fit the language of the respondents, that deal with a single issue at a time, that are ordered to maximize interviewee motivation, interest, and willingness to respond,

and that reflect the results of a careful pretest and the resolution of problems most likely to be encountered in the large-scale data collection.

However, even when the interviewer is armed with a near-perfect interview schedule, the quality of the data obtained ultimately depends on his or her training and performance. Interviewers are the "foot soldiers" in survey research; consequently, most survey research organizations pay close attention to their selection and training. Let us briefly highlight some of the main issues involved in training interviewers.

Selecting Interviewers

Given the importance of the interviewer to the entire research process, initial selection of potential interviewers is a critical decision point in any project. Anyone's interviewing skills can probably be improved in a well-developed training program, but some people have the ability to establish rapport quickly and are better than others at, for example, conducting an interview in a conversation-like, supportive manner. In other words, most training programs are not sufficient to make good interviewers out of poor ones; the social skills that accompany successful interviewing are generally part of the personality by the time one reaches young adulthood. Backstrom and Hursh (1963: 134) list the following characteristics of good interviewers:

1. Completely honest in their work.
2. Reliable and conscientious.
3. Utterly objective in their manner of asking questions.
4. Faithful and neutral in recording answers.
5. Willing to write answers fully and legibly.
6. Interested in people; understanding.
7. Able to inspire people's confidence and put them at ease.
8. Inconspicuously but neatly dressed.

They further indicate that good interviewers are willing to study all questions thoroughly until they know what the items mean and they will regularly reread their instructions to pick up anything that they might have missed or to avoid getting into bad habits that need to be corrected. The good interviewer exhibits an attitude of neutrality (in most instances, they are "merely to soak up information like a sponge without giving any of it back") so that their own attitudes and actions do not affect the respondent's answers. The good interviewer is conversational and friendly, but also impartial, not allowing his or her opinion to influence the course of the interview. Backstrom and Hursh (1963: 137) conclude: "The object of the survey is to get the honest, uninfluenced opinion of each individual interviewed. You (the interviewer) are merely the medium through which the opinion is conveyed. Nothing of you should be in the interview results."

Some other characteristics of interviewers may also be relevant. For example, one must always consider the potential effect of an interviewer's ethnicity, gender, age, or other personal characteristics on the people to be interviewed.

It is not always possible to find ideal interviewers having the characteristics outlined above. In any case, careful attention to the initial selection of interviewers is time well spent.

Training Interviewers

Innumerable lists of "do's" and "don'ts" for interviewers could be presented. The following list has been frequently used by the authors in their own training of interviewers. Each of these points should be periodically reviewed during interim training sessions or staff meetings:

1. Please keep this list for future reference. Additions will be made throughout the project as the need arises.
2. *Always* glance through the interview schedule when it is picked up in the office to see if all pages are present, in sequence, and legible.
3. *Always* leaf through the interview schedule (page by page) *before terminating the interview* to see if you have overlooked a question or if you are unclear as to an interviewer's response(s). In those very rare circumstances where the respondent seems reluctant to divulge certain information and when probing seems unwise for the time being, clearly mark that item so that it will be drawn to your attention in your final review of the schedule. Use tact, however, in questioning a respondent. Often rapport will have developed during the interview so that a question which the respondent initially refused to answer is answered without objection when asked again at the end of the interview.
4. It is important to most studies—especially during the pilot phase—that we weed out those items that seem not to "work." When an item is suspect, in your opinion, identify it on the schedule so that it can be recalled for discussion in staff meetings. *On the other hand*, do not allow your suspicions to affect the way in which you present the item to the interviewee.
5. On recording answers to precoded questions: When a respondent persistently responds in a way that seems not to fit the categories (or the response is qualified in some way that we had not anticipated), record all relevant comments in the margins so that coders will be able to interpret the exceptions. This practice is especially important in pilot studies, since it may suggest a need for revision of the interview schedule.
6. With open-ended items it is essential that you record all relevant material. If there is some question as to the relevance of some statements, record them nonetheless. *When in doubt, write it down.* It is inevitable that you will develop abbreviations in recording replies. Penmanship may also suffer in the rush of keeping pace with the respondent. It is therefore vital that you review your schedule and "translate" illegible script and obscure abbreviations so that coding is simplified. It is best that this be done immediately after the interview. At the same time you will want to add to the instrument your own impressions of the respondent and the interview situation.

7. Insofar as possible, do *not* read items from the schedule as one might if they were quoted from a book. That is, the interview should resemble a conversation. On the other hand, it is important that you present the items to each respondent exactly as worded on the schedule. Your facility at presenting the items in a conversational way will grow with experience.
8. Do not put words into the mouth of the respondent. When an interviewee seems unable to articulate his or her thoughts, be cautious in probing so that you do not bias the response. Remember that a respondent is affected by your reactions (words, gestures, and so forth). Expressions that imply value judgments are to be avoided. At the same time, a limited amount of interviewer response is sometimes effective and necessary in establishing rapport and in "inspiring" the respondent's powers of recall. Such interviewer response should be neutral but supportive. Do not express your own ethical, moral, political, or other opinions.
9. Do not assume too much. When the interviewee has not been clear in a reply, *never* infer or assume that you have the missing data. Continue to probe. The fact that you have interviewed the person for an hour does not necessarily mean that you know what he or she means or would respond. People are often amazingly inconsistent.
10. When inconsistencies are observed in responses, it is necessary to question the respondent. Be tactful so that you do not give the respondent a feeling of being "trapped." Use phrases such as, "I'm not certain I understand . . .," "I think I'm unclear since you mentioned that . . .," and so forth. Do not appear to "interrogate" the interviewee.
11. Whenever you are confronted with a situation that is in some way out of the ordinary, jot it down so that it can be discussed in staff meetings. If you have developed what you believe to be a fruitful and effective approach to any interview problem, it should be discussed and shared with the field supervisor and interviewing staff.
12. No question is too personal—if it is asked properly.
13. Keep the interview moving along at a comfortable pace, but do not sacrifice quality. This pace, of course, will be determined by the respondent's powers of recollection.

Finally, any training program must prepare interviewers to respond to a variety of common questions asked frequently in any survey. Backstrom and Hursh (1963: 143–144) have a list of some stock answers to frequent questions. We repeat their list below:

WHAT YOU SHOULD SAY . . .

1. **If respondent asks, "Who is doing this survey?":** "This survey is being conducted by the Research Division of _____. We are trying to get some ideas about what people think about current issues in _____."
2. **If respondent presses for a better answer on auspices:** "Well . . . I'm a professional interviewer. The people in charge of this survey are at the Research Division at _____. They'd be glad to explain the survey to you. Would you like their phone number so you could call them?" (If "Yes," give trouble number.)

3. **If respondent wonders why he is being interviewed or suggests interviewing someone else:** "You were selected completely *by chance* according to procedures worked out by my office. So *your* opinions are important and interviewing someone else wouldn't be as *good*."
4. **If respondent says he doesn't have the time to be interviewed:** "The questions won't take long. You can go right on with your work and I'll just run through these items." (Begin questioning immediately.)
5. **If respondent insists he is too busy:** "What would be a better time soon for me to come back? I'll note down an appointment that would be more convenient for you."
6. **If respondent says he doesn't know enough to give good answers:** "In this survey, it's *not* what you know that counts. Rather, it's what you happen to think about various topics that is important."
7. **If respondent is afraid to answer some questions or asks, "What are you going to do with these answers?" or "Why do you want to know that?"** "Well . . . many people are being asked these same questions, of course, and what you say is confidential. We are interested in these questions only to see what a *lot* of people in _____ generally are thinking about."
8. **If respondent resents questions that talk down to him:** "The people in my office made up these questions, and we are instructed to read each one just as it is written."
9. **If respondent is annoyed and just plain refuses to answer a question:** "Of course, you don't have to answer any question you'd prefer not to. I'm only trying to get your opinion because our study is more accurate that way." Then if respondent still refuses, don't comment, just go on quickly to the next question. Mark the item "Refused."

This list of stock answers will not solve all problems an interviewer will face. However, they should be a standard part of his or her repertoire of "tools of the trade."

We have not outlined a complete interviewer training program but rather have tried to present some basic parts of any training program. Much of the specific content of such training programs is unique to each project. In all cases, however, adequate attention must be given to the teaching of interviewers if the data collected are to be valid and reliable.

Approaching the Respondent

When the interviewer has the final version of the interview schedule, the next important consideration is to learn how to approach and interview the potential respondent. Most training manuals emphasize the point that we have stressed earlier, that the primary task at this point is to motivate the respondent to participate. The task of motivation is usually accomplished by stressing several points in an introductory statement.

First, the respondent must be convinced that the study is important and worth the time he or she is being asked to spend in talking to the interviewer. Some studies carry a built-in interest for the respondent. For example, the "expert informants" typically interviewed in social impact

studies (see Chapter 12) are selected because they are thought to know about how a proposed project might impact on themselves and their community. Such respondents usually have a high level of interest in the project and are most willing to be interviewed. Other studies appeal to group memberships or other salient personal characteristics of the respondent. For example, in their nationwide study of Lutherans, Strommen et al. (1972) emphasized to their respondents that they were being asked to participate in the biggest and most comprehensive study of Lutherans that had ever been made and that the study would have great significance for all members of their church. Unless a respondent can appreciate the significance of a study, he or she is not likely to agree to participate.

Second, in terms of sources of intrinsic motivation that we discussed earlier, the respondent must feel that the interview will be a pleasant and satisfying, or at least not an unpleasant experience. This means that interviewers must present themselves as pleasant and easy to talk with and must define the interview as an opportunity for the respondent to express important personal views and feelings.

Finally, the interviewer must learn to deal with unanticipated barriers to the interview. Such barriers might include questions the respondent has about the identity of the interviewer, the legitimacy of the organization conducting the study, concerns about why he or she was selected to participate in the study, fears about anonymity or how the data are to be used, and so on.

The Institute for Social Research at the University of Michigan (1969) provided its interviewers with a list of several specific ways an interviewer might respond to the above barriers.

1. Tell the respondent who you are and whom you represent. In other words, interviewer identity and sponsorship must be made clear. The interviewer will need to carry some form of identification and should always have a telephone number that the respondent can call to verify the interviewer's identity and sponsorship.
2. Tell the respondent what you are doing in a way that will stimulate his or her interest. Again, this is basically the task of making the respondent's participation in the study personally meaningful.
3. Tell the respondent how he or she was chosen. This is one of the most common questions in survey research—respondents want to know why they were selected to participate in the study. If random sampling techniques have been used, this should be briefly explained to the respondent. The explanation should emphasize why it is important that the respondent participate and why a substitute would not be as good.
4. Finally, the interviewer must, through both the presentation of self and the introduction of the study, create a relationship of confidence and understanding with the respondent.

Other factors should be considered in planning the all-important approach to the potential respondent. Some researchers recommend making

a telephone appointment with the subject before going into the field. Others inform potential subjects about the study by mail and then "take their chances" when they try a door approach without a definite appointment. Others prefer the person-to-person description of the study at the respondent's doorstep.

VIII. OTHER VARIATIONS: THE TELEPHONE INTERVIEW

Large-scale interview surveys are costly and time-consuming, especially when respondents are spread over a large geographic area and when several call-backs are required to complete the necessary interviews. The high cost of data collection by personal interview has led researchers to seek alternatives, and in recent years a popular approach has been interviewing by telephone. Perhaps the best way to describe telephone interviewing as a research tool is to compare it with traditional personal interviewing in terms of advantages and disadvantages.

Advantages of Telephone Interviews

The primary advantages associated with telephone interviewing are its lower cost, more manageable sampling procedures, convenience to interviewers, and accessibility of interviewers to supervision and management.

Cost The costs of salaries and travel for interviewers continue to increase. The average cost per interview in a large sample survey may be over \$100. Telephone interviewing is much less expensive because all of the interviews are conducted from one or a few centrally located research facilities. Often a large bank of telephone stations is set up in a single room and interviewers work in shifts calling and interviewing.

Sampling As was noted in a previous chapter, area sampling is often difficult and expensive. However, most households in the United States and in other industrialized countries have telephones. In the United States the number of telephones approaches 95 percent of all households. Accordingly, lists of telephone numbers provide an excellent sampling frame for survey research. Further, techniques have been developed to deal with the possibility of unlisted numbers or persons who have recently arrived in an area and whose names are not listed in the published telephone books. A procedure generally referred to as random digit dialing for telephone surveys (see Klecka and Tuchfarber, 1978) allows random access to all working numbers; all exchanges in a vicinity are identified and then the exchange prefixes are randomly dialed, followed by a random number

between 0001 and 9999. This procedure gives everyone in a given exchange area who has a telephone an equal possibility of being selected for inclusion in the survey.

The development of a procedure for the random generation of sample telephone numbers combined with other methodological refinements of telephone interviewing techniques has been viewed by some as the most important advancement in survey research since the introduction of area probability sampling and the use of computers for data storage and analysis (Groves and Kahn, 1979; see also Dillman, 1978).

Convenience and access Calling someone on the telephone is much more convenient than making a personal visit. One does not have to worry about finding addresses, finding particular respondents within the household, coordinating travel schedules, and so on. It also provides some advantages in gaining access to respondents. Residents of large cities may be hesitant to invite unknown people into their homes or to answer their questions but they are often willing to participate in the less threatening telephone interview.

Supervision Because all of the telephone interviewers operate from one or a few central locations, supervision is much simplified, and thus consistency and overall quality control are improved. In addition, supervisors can work out schedules so that calls are made during various times of the day or evening and interviewers most adept at handling problems can be assigned the most difficult cases.

Disadvantages of Telephone Interviews

There are also some disadvantages to interviewing by telephone. Among the most important of these are the limitations on the length of the interviews, some constraints on topics that may be treated, and the ease with which respondents may terminate an interview. A further disadvantage of not confronting the respondent in person is not being able to establish rapport and a temporary sense of cooperation sufficient to minimize evasive or incomplete responses. The absence of eye contact is another important drawback. Telephone rapport is rarely as strong as rapport in contexts of personal confrontation.

Length of interview and nature of materials to be covered It is difficult to conduct a telephone interview that lasts more than 15 to 20 minutes and many respondents will tire before this point. In addition, telephone interviewing generally produces much less information than would be obtainable via personal interview. Some research suggests that if respondents are contacted by letter prior to the telephone call, more complete infor-

mation can usually be obtained (Tremblay, 1977). Questions in telephone interviews must be brief and simple. The interviewer has few opportunities to establish rapport or to give a convincing account of the importance of the study. Other types of information, such as the attitude of the respondent during the survey, the nature of his or her surroundings, the presence of other persons, and similar aspects of the setting at the respondent's end are simply not available to the interviewer. Finally, people are often more reticent about personal information such as income in a telephone interview than they would be in a traditional interview.

Problems of interview termination Respondents in telephone surveys can quickly terminate the interview at any point merely by hanging up the phone. This option is taken often enough that researchers who collect data by telephone have higher portions of missing data on individual items. It is much more difficult for a respondent in a face-to-face situation to terminate an interview before it is completed. Results from telephone surveys exhibit more missing data on family income and more evasiveness and response bias on attitudinal questions (Jordan, Marcus, and Reeder, 1980).

Despite these limitations, because face-to-face interviewing in survey research is so expensive, we can anticipate a growing reliance on telephone interviewing, particularly when only moderate detail is required (Goldi and Pritchard, 1981). However, many topics are simply not appropriate for telephone interviewing, and so the traditional face-to-face interviews will not disappear but will come to represent a smaller percentage of data collection efforts than in the recent past.

IX. COLLECTING DATA BY INTERVIEWING: AN ILLUSTRATIVE EXAMPLE

As was noted earlier, there are now many public and private research organizations that regularly conduct large-scale interview and questionnaire surveys. Some of the most widely known and visible examples are the American Institute of Public Opinion (Gallup Poll), the Roper Public Opinion Research Center, the National Opinion Research Center (NORC), the Harris Research Organization, and the Michigan Survey Research Center. These and many other organizations are involved in frequent assessments of the attitudes, beliefs, and behaviors of the national and various subcultural or regional populations.

Students interested in survey research can refer to these and other professional research organizations for illustrative copies of the interview schedules, standard coding instructions, and handbooks for interviewer training and management.

As a brief example, we will review the procedures used by the Na-

tional Opinion Research Center in one of their long-term projects. NORC has been conducting General Social Surveys on an annual basis since 1972. Each survey is conducted with an independently drawn sample of English-speaking persons 18 years of age or over, living in noninstitutionalized arrangements within the continental United States (NORC, 1982). Interviews are conducted by highly trained interviewers and typically run for about one hour.

Interviewers for the NORC General Social Surveys are hired and trained by area supervisors. Interviewers conduct practice interviews which are evaluated by NORC, and performance must reach an acceptable standard before the interviewers are allowed to begin their formal fieldwork. Completed interviews are returned to NORC headquarters, where they are edited for completeness and accuracy. Twenty percent of the interviews are validated by someone from NORC recontacting the person with whom the interview was conducted. The interview schedules are then coded, the data entered on computer tape, and the data made ready for analysis.

Items in the General Social Surveys are generally drawn from previous NORC studies and from those conducted by other national polling organizations in order to facilitate time trend studies and the replication of earlier findings. In order to maximize reliability and validity, all items used in the General Social Survey schedules have been tested previously. There are three general types of questions: permanent ones that appear in each annual survey; rotating questions that appear in two of every three surveys; and "occasional" questions that appear in one year's instrument and are replaced the next year (NORC, 1982).

Virtually all of the items in the NORC surveys are structured in a closed format. For example, a question on marital happiness asks: "Taking all things together, how would you describe your marriage? Would you say that your marriage is very happy, pretty happy, or not too happy?" Similarly, a question inquiring about health asks: "Would you say your own health, in general, is excellent, good, fair, or poor?" Interviewers are provided clear directions for handling problems that might emerge. For example, in the instruction accompanying the question, "What race do you consider yourself to be?" the interviewer is instructed to code the question without asking only if there is absolutely no doubt about the racial identity of the respondent. Otherwise, they are instructed to ask about racial identity and are provided with the following rule: "A person should be classified as 'Other' only if he is American Indian, Japanese, Chinese, Filipino, Asian Indian, Korean, Polynesian, Indonesian, Hawaiian, Aleut, or Eskimo. A person is classified as 'Black' only if he is American Negro; or if he is African, West Indian, or Puerto Rican who appears to be black. All other persons are classified as 'White.'" This includes Mexicans, Spaniards, and also Africans, West Indians, or Puerto Ricans who appear to be white" (NORC, 1982: 229-230).

Interviewers are also given clear directions about probe questions. For example, in response to the question, "How often do you attend religious services?" the interviewer is instructed to use a standard series of categories as probes, if probing is necessary.

Each step in the NORC procedure has been developed through prior experimentation with alternative approaches. The outcome is a highly sophisticated method of collecting data from representative samples of the American public, and the data are available to the public or to individual researchers at minimal cost.

X. SUMMARY

Survey research is the most widely used method of data collection in the social sciences. Surveys usually involve interviewing or the use of some form of self-administered questionnaire. Interviews are a type of two-person conversation that is initiated by the researcher for the purpose of obtaining research-relevant information.

Despite its widespread use, interviewing has some serious drawbacks as a method of data collection. Among the most important sources of error in interview data are: (1) errors that result from purposeful intent by the respondent to deceive or mislead, (2) problems associated with the temporary role of respondent, (3) the mental-psychological state of the respondent, and (4) involuntary error from such things as forgetfulness or lack of information. Other sources of bias in interviewing stem from the personal attitudes of the interviewer, or visible characteristics such as age, gender, and ethnicity that may suggest to the respondent the interviewer's social status or group membership. The personal attributes of the interviewer are among the cues that the respondent will use to attribute attitudes and motives to the interviewer, and these attributes, in turn, will influence response to interview items.

Interviews fall along a continuum ranging from those that are highly structured on the one end to those exhibiting virtually no structure on the other. Highly structured instruments usually contain a series of very specific questions, and sets of predetermined response categories. Unstructured interviews are designed to help the interviewer explore a variety of preselected topics but not to bind him or her to asking specific questions in any preestablished format or sequence. The more structured interview designs are usually selected for those situations where the dimensions of what one wishes to study are most clearly understood. When such dimensions are less clearly established, the researcher may profit from a more open format, perhaps doing a series of in-depth interviews to gain a better picture of the scope and characteristics of an unknown population.

One of the most important tasks facing the researcher who plans to collect data by interview is formulating the questions that are to be asked. Good questions provide the information needed to test one's research hypotheses, and at the same time must be asked in ways that motivate respondents to answer fully and truthfully. This means that questions must be understandable and be properly interpreted by the respondent. A good question must also elicit personal interest and a desire to provide the information called for. Typical problems to avoid in formulating good items are double-barreled questions (two or more questions woven together in a single item), questions that are not simple enough so that relatively uneducated respondents can understand, and questions that personally seem threatening and embarrassing (particularly if sufficient interest and rapport have not already been clearly established); also, "fail-safe" procedures should be established for probing or pursuing a particular question in sufficient detail so that the necessary information can be obtained even from taciturn respondents.

Because of the many potential problems in survey research, the researcher must devote much effort to advance preparation, identifying potential difficulties, and learning to avoid them or be minimally affected by them. Before beginning the task of writing questions for an interview, the researcher should learn as much as possible about the topic to be studied and about the people who will be respondents. The former information is necessary if the researcher is to ask the right questions; the latter is necessary if the right questions are to be asked in the right way, thereby producing a maximum amount of relevant data in return for the costs of the research to both the research sponsor and professional team and also to the respondents whose lives are affected by the obtrusive nature of the interview.

An important issue in all types of surveys is that of whether to use open or closed questions. With an open format, the respondent answers a question in his or her own words. Closed category items require the respondent to select from among a series of alternatives provided by the researcher. Whether one chooses an open or closed format for questions will usually be determined by the following factors: (1) interview objectives, (2) respondent information level, (3) the structure of respondent opinion, (4) respondent motivation to communicate, and (5) the researcher's knowledge about the respondents and their social context.

Because the interviewer is the primary data-collector in interview research, there has been much work on developing techniques for training and supervising interviewers. When a well-framed, appropriate set of questions is placed in the hands of properly trained, motivated, and supervised interviewers, the quality of the resulting data is usually very high. However, a breakdown at any of these levels greatly decreases the quality of

the data obtained and therefore the degree of confidence researchers can place in their findings—and frequently too little time is devoted to instrument construction, pretesting, and interviewer training.

Recently there has been a rapid growth in the use of telephones for conducting interviews and the use of telephone books or random digit dialing techniques to construct samples and sampling universes. Telephone interviews offer several advantages, including greatly reduced costs. However, they also have some important limitations, including constraints on the length and topic of the interview and the ease with which respondents can cut off an interview or refuse to be interviewed at all.

REFERENCES

- BACKSTROM, CHARLES H., and GERALD D. HURSH, 1963. *Survey Research*. Chicago: Northwestern University Press.
- BAHR, HOWARD M., and KATHLEEN C. HOUTS, 1971. "Can You Trust a Homeless Man? A Comparison of Official Records and Interview Responses by Bowery Men." *Public Opinion Quarterly* 35 (Fall): 374-382.
- BALE, RICHARD N., 1979. "The Validity and Reliability of Self-Reported Data from Heroin Addicts: Mailed Questionnaires Compared with Face-to-Face Interviews." *The International Journal of the Addictions* 14, No. 7: 993-1000.
- BALL, JOHN C., 1967. "The Reliability and Validity of Interview Data Obtained from 59 Narcotic Drug Addicts." *American Journal of Sociology* 72 (May): 650-654.
- BECKER, HOWARD S., and BLANCHE GEER, 1957. "Participant Observation and Interviewing: A Comparison." *Human Organization* 16 (Fall): 28-32.
- BENNEY, MARK, and EVERETT C. HUGHES, 1956. "Of Sociology and the Interview." *American Journal of Sociology* 62 (July): 137-142.
- CAMPBELL, BRUCE A., 1981. "Race-of-Interviewer Effects Among Southern Adolescents." *Public Opinion Quarterly* 45 (Summer): 231-244.
- CANNELL, CHARLES F., and ROBERT L. KAHN, 1968. "Interviewing." Pp. 526-595 in Gardner Lindzey and Elliot Aronson (eds.), *The Handbook of Social Psychology*, Vol. 2. Reading, Mass.: Addison-Wesley.
- CANNELL, CHARLES F., LOIS OKSENBERG, and JEAN M. CONVERSE, 1977. "Striving for Response Accuracy: Experiments in New Interviewing Techniques." *Journal of Marketing Research* 14 (August): 306-315.
- DENZIN, NORMAN K., 1970. *The Research Act*. Chicago: Aldine.
- DILLMAN, DON A., 1978. *Mail and Telephone Surveys: The Total Design Method*. New York: John Wiley.
- GOLDI, JUDITH, and JACKI PRITCHARD, 1981. "Interview Methodology—Comparison of Three Types of Interview: One to One, Group and Telephone Interviews." *Aslib. Proceedings* 33 (February): 62-66.
- GROVES, ROBERT M., and ROBERT L. KAHN, 1979. *Surveys by Telephone*. New York: Academic Press.
- HANSON, JIM, 1979. "On How to Live with Survey Research." *Contemporary Sociology* 8 (September): 677-679.
- INSTITUTE FOR SOCIAL RESEARCH, 1969. *Interviewer's Manual*. Institute for

- Social Research, Survey Research Center, Ann Arbor, Mich.: University of Michigan.
- JORDAN, LAWRENCE A., ALFRED MARCUS, and LEO REEDER, 1980. "Response Styles in Telephone and Household Interviewing: A Field Experiment." *Public Opinion Quarterly* 44 (Summer): 210-222.
- KAHN, ROBERT L., and CHARLES F. CANNELL, 1957. *The Dynamics of Interviewing*. New York: John Wiley.
- KINSEY, A. C., W. B. POMEROY, and C. E. MARTIN, 1948. *Sexual Behavior in the Human Male*. Philadelphia: W. B. Saunders.
- KLECKA, WILLIAM R., and ALFRED J. TUCHFARBER, 1978. "Random Digit Dialing: A Comparison to Personal Survey." *Public Opinion Quarterly* 42 (Spring): 105-114.
- MACCOBY, ELEANOR, and N. MACCOBY, 1954. "The Interview: A Tool of Social Science." Pp. 449-487 in G. Lindzey (ed.), *Handbook of Social Psychology*, Vol. 1. Cambridge, Mass.: Addison-Wesley.
- NATIONAL OPINION RESEARCH CENTER, 1982. *General Social Surveys, 1972-1982: Cumulative Codebook*. National Opinion Research Center, University of Chicago.
- NELSON, HAL, BRUCE E. BECHTOL, and THOMAS W. JOHNSON, 1977. "Interviewing and Observational Techniques in Social Science Fieldwork." *Journal of Geography* 76 (March): 94-99.
- RICE, S. A., 1929. "Contagious Bias in the Interview: A Methodological Note." *American Journal of Sociology* 35 (November): 420-423.
- SCHUMAN, HOWARD, and J. CONVERSE, 1971. "The Effects of Black and White Interviewers on Black Responses in 1968." *Public Opinion Quarterly* 35 (Spring): 44-68.
- SCHUMAN, HOWARD, and STANLEY PRESSER, 1977. "Question Wording as an Independent Variable in Survey Analysis." *Sociological Methods and Research* 6, No. 2: 151-170.
- 1979. "The Open and Closed Question." *American Sociological Review* 44 (October): 692-712.
- 1981. *Questions and Answers in Attitude Surveys: Experiments on the Effects of Question Form, Wording and Context*. New York: Academic Press.
- STROMMEN, MERTON, MILO BREKKE, RALPH UNDERWAGER, and ART JOHNSON, 1972. *A Study of Generations: Report of a Two-Year Study of 5,000 Lutherans Between the Ages of 15-65, Their Beliefs, Values, Attitudes, Behavior*. Minneapolis: Augsburg.
- TREMBLAY, KENNETH R., 1977. "Interviews by Telephone and in Person—Quality of Responses and Field Performance." *Public Opinion Quarterly* 41 (Spring): 117-118.
- VIDICH, ARTHUR, and JOSEPH BENSMAN, 1954. "The Validity of Field Data." *Human Organization* 13 (Spring): 20-27.
- WILLIAMS, J. ALLEN, 1968. "Interviewer Role Performance: A Further Note on Bias in the Information Interview." *Public Opinion Quarterly* 32 (Summer): 287-294.

Survey Research: Questionnaire Studies

six

I. Introduction

**II. Advantages of Survey by
Questionnaire**

**III. Disadvantages of Survey by
Questionnaire**

**IV. Doing a Survey by Ques-
tionnaire**

Statement of the problem

*Identification of population and
sample selection*

Questionnaire construction

*Administering the question-
naire*

Analysis and report

**V. An Illustrative Example:
Equality of Educational Op-
portunity**

VI. Summary

I. INTRODUCTION

In **questionnaire** surveys, a respondent fills out and returns to the researcher a self-administered "interview," in which the questions and instructions are complete and understandable enough that the respondent can act as his or her own "interviewer." The questionnaires may be hand-delivered to the respondent's home or office or be distributed at locations such as shopping malls or movie theaters. They may be distributed to groups of respondents such as university students, military recruits, church congregations, airline passengers, or attenders of civic society meetings. In such settings the questionnaires are distributed to group members, but each fills out the questionnaire privately. Perhaps the most popular strategy for the distribution of questionnaires is to mail them, along with a request that the completed questionnaire be returned by mail to the researcher.

Because the questionnaire is self-administered, it depends on a written or spoken (in group settings) appeal to motivate respondents to participate in the study, and upon instructions and questions that interest respondents to "carry them along" through the task of completing and returning the form. There is no interviewer to cajole or build a rapport, to clarify items, or to reward participants with friendly statements.

The nature of questionnaire studies is best illustrated by example. Let us begin with a description of a questionnaire survey on religiosity and

delinquency conducted by two of the authors (Albrecht, Chadwick, and Alcorn, 1977). The questionnaire contained items asking teenagers how often in the past two years they had smoked cigarettes, consumed beer or hard liquor, smoked marijuana, taken LSD or similar drugs, petted on a date, had premarital sexual relations, stolen things costing more than two dollars, shoplifted, or started fights. Religious behavior was measured by asking the young people how often they attended church services. Religious attitudes were reflected in a four-item religiosity scale focusing on belief in the existence of God, the existence of Jesus Christ, the existence of the devil, and the validity of the Bible. Additional questions probed the strength of the respondents' relationships with their parents, their participation in family activities, and the strength of their ties to their peers.

Data were collected from families residing in six congregations of the Mormon church. Two of the congregations were located in small rural towns in southern Idaho, two were in a city in central Utah, and there were two in the suburbs of Los Angeles. The homes of the rural Idaho congregation were so scattered that we opted to contact them by mail. The **personal delivery technique** was used in the Utah and California settings. Lists of the families in the four Utah and California congregations, including the names and ages of children, were obtained from local church leaders. Packets were prepared containing questionnaires for each child twelve years of age and older, and one for each parent. An envelope, addressed to the research office, was also provided for each family member. The packets were delivered to the selected homes by a research assistant, who explained the project to one of the parents. A written introduction to the project and instructions on how to fill out the questionnaire were attached to the packet. The introduction explained that the objective of the project was to assess the relationship between religion and delinquency, and it stressed the importance of each family's participation. The instructions went on to ask that each respondent complete the questionnaire alone, then emphasized the need for accurate information and promised that all responses were confidential. The research assistant returned to the households a few days later to pick up the completed questionnaires. If family members had not finished the task, the research assistant asked that they do so as soon as possible and returned again the following day. A third call-back was made if necessary. Persons who had not completed the questionnaire after three personal call-backs were counted as refusals.

The six congregations contained 409 teenagers. Completed questionnaires were obtained from 224 youths, a 60 percent response rate. There were few refusals. Most of the nonrespondents were members of families on vacations or of families experiencing problems such as serious illness.

The results revealed that religious participation and, to a lesser degree, religious attitudes, were related to lower rates of delinquent behavior. Moreover, by combining information on religious activity and attitudes, the

quality of family relationships, and the nature of peer relationships, much of the self-reported "deviance" could be explained.

II. ADVANTAGES OF SURVEY BY QUESTIONNAIRE

The major advantage of the questionnaire survey is economy. It yields the maximum number of "facts," or bits of data, per research dollar. However, the economy is gained via built-in limitations on the depth of the data obtained. Costs for interviews vary depending on the nature of the sample and the interview length, but even under optimal conditions (the sample being concentrated in a limited geographical area, the interview being brief, and few call-backs being necessary), interviews cost at least \$25 each and interviews in contemporary research usually range between \$50 and \$150. On the other hand, questionnaire surveys obtain completed questionnaires for about \$3 each. Dillman (1978: 69) reported that when first-class postage was 8¢, the cost of a completed questionnaire in a mail survey ranged from \$1.60 to \$2.84. Raises in postage have increased the cost another 30¢ to 50¢ per questionnaire, but the cost is still much less than that of an interview. Most times a budget of under \$3,000 is sufficient to produce completed questionnaires from a sample of 1,000 respondents.

A second advantage of survey by questionnaire is that the respondent may consult with others, review records, think about a question before answering, and interrupt the process of completing the instrument if necessary. In an interview setting such time-consuming or disruptive actions are usually inappropriate.

Some researchers argue that the questionnaire survey is a useful way to obtain information about sensitive topics. It is suggested that to report how often one exhibits deviant or disapproved behavior is easier for the respondent working on a questionnaire than it would be in face-to-face interaction with an interviewer. If the respondent is convinced that the questionnaire is anonymous, he or she can freely report attitudes and behaviors without embarrassment or fear of reprisal (Sudman and Bradburn, 1982).

Wiseman (1972) compared responses to questions on socially sensitive issues in **face-to-face interviews**, **telephone interviews**, and **mail questionnaires**. A sample of 640 residents of a suburb of Boston were selected and divided into three subsamples, one for each of the three data-collection techniques. Nine questions on controversial issues included queries about an all volunteer army, legalizing marijuana, making birth control readily available to unmarried people, legalizing abortion, and lowering the drinking age to 18. For seven of the nine issues, the responses obtained by the three different techniques were very similar. However, the comparisons also revealed that support for birth control and abortion was higher in the

questionnaires than in the interviews. For example, 89 percent of the questionnaire respondents favored legalized abortion as compared to 70 percent of those interviewed face-to-face and 67 percent interviewed over the telephone. Wiseman interpreted these findings as evidence that the questionnaire was a more sensitive instrument in that questionnaire data were less biased toward socially desirable responses than were the data obtained from interviews either by telephone or in person.

Dillman concluded from an extensive review of the research literature that mail questionnaires produce fewer responses biased in the direction of social desirability than do interviews. In his words, "The greatest advantage of face-to-face interviews—the physical presence of the interviewer—may at times be [their] greatest drawback" (Dillman, 1978: 63).

III. DISADVANTAGES OF SURVEY BY QUESTIONNAIRE

One limitation of questionnaire studies is that questionnaires must be relatively brief or most respondents will not take the time to complete them. Personal interviews may last as long as six or seven hours, and routinely they take more than an hour. Certainly interviews, when properly conducted, have a longer potential duration than questionnaires. Under some circumstances lengthy questionnaires can be administered. We noted in the previous chapter that Strommen and his associates (1972) successfully fielded a three-hour questionnaire.

Dillman reviewed the literature on the effect of questionnaire length on response rate and quality of response and tentatively concluded that 11 pages, or 125 questions, was the limit beyond which the response rate fell off (Dillman, 1978: 55). It is suspected that 11 to 12 pages is the point at which the quality of responses also declines. Near the end of a long questionnaire respondents are tired and tend to glance at the questions and hurriedly answer, sometimes inappropriately, in order to finish as quickly as possible.

Another limitation of the questionnaire is that the researcher cannot **probe** or follow up on interesting leads. An interviewer can pursue an unanticipated but insightful comment, and this ability to probe new ideas is important, especially in exploratory research. The limitation on follow-up of unanticipated responses can be compensated for to a degree by including opportunities for respondents to express their feelings in an unstructured way. Open-ended questions and instructions to comment about the structured questions make it possible for respondents to treat some topics in detail.

Advocates of questionnaire surveys argue that sensitive topics can be studied with less bias by questionnaire than by interview. On the other hand, some researchers say that interviews produce better data. Mangione, Hingson, and Barnett (1982) asked the same questions in a face-to-face

interview, a telephone interview, and a questionnaire. Their data were collected from three identical samples in the Boston metropolitan area during 1977. Among the items were some having high components of "social desirability," including questions on how happy, depressed, involved, lonely, excited, or bored the respondents were, and also questions about their drinking patterns. The researchers found that the interview, either face-to-face or by telephone, generated more extreme answers at both ends of the happiness, depression, etc. continua than did the questionnaire. The differences between questionnaires and interviews in reported drinking behavior was modest, but the respondents who admitted the heaviest drinking were those questioned by interview. Contrary to the findings reported above by Wiseman, the more deviant (or extreme) behaviors reported in the interviews were interpreted as evidence that interviews obtain more valid responses to socially sensitive questions. However, whatever technique is used, it is hard to be sure of obtaining truthful information about deviant or disapproved behaviors. Fortunately, the use of both questionnaires and interviews in the same study has shown that the limitations of each method may be offset by the strengths of the other.

Another potential problem with questionnaire studies is the risk that someone other than the selected respondent will complete the questionnaire. A researcher may ask the husband to participate, but there is no guarantee that the potential respondent's wife will not try to be "cooperative" and complete the questionnaire for her husband. She may even think she knows how her husband would respond. This problem of substitution of respondent can be handled by stressing in the introductory communications and again in the instructions to the survey that the respondent has been selected as part of a scientific sample, that his or her responses are critical, and that substitutions are not permitted.

A final limitation in questionnaire surveys is that a respondent can change his or her answers. In an interview study the respondent may be misled about the real intent of the study until the last few questions. In this case, interviewers will not change earlier answers, whereas with a questionnaire respondents can make any alterations they wish. Thus a questionnaire should not be used in a study where some form of deception is necessary and where later questions reveal the researcher's intent. Respondents will likely go back and change the answers initially entered in the first part of the questionnaire.

IV. DOING A SURVEY BY QUESTIONNAIRE

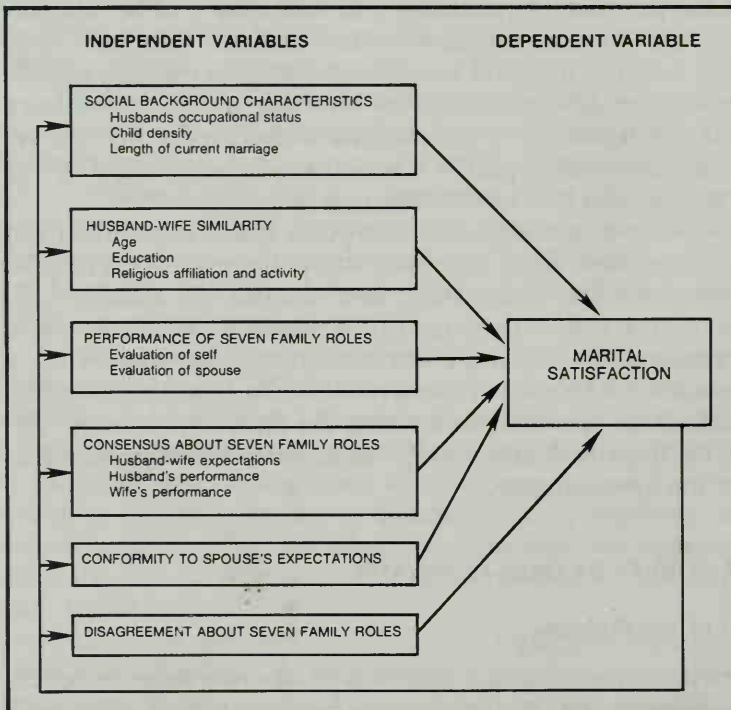
Statement of the Problem

The objective of questionnaire studies is to describe some behavior, attitude, or feeling or to test the relationship between two or more such characteristics or variables. For example, the researcher may be interested

in describing how people feel about legalized abortion, how many women have had an abortion, how many families buy a certain food product, which candidate people intend to vote for, workers' attitudes about their jobs, how often workers have missed work or taken company goods for their personal use, or how often parents discipline their children. The researcher may want to test a relationship between two or more variables such as the association between religious affiliation and support for abortion or the relationship between levels of salary and job satisfaction. We will illustrate the steps of questionnaire study with an example in which the authors (Chadwick, Albrecht, and Kunz, 1976) tested the relationship between social background characteristics of husbands and wives, their reports of who performed various family roles, the degree of consensus between husbands and wives about who should perform the roles, their conformity to each other's expectations, and their feelings of marital satisfaction (see Illustration 6.1).

The arrows from the independent variables to the dependent variable identify an association between them and imply that they are causal relationships. It must be stressed that survey data will not document a causal relationship and that, in fact, the relationship is probably circular. In other

ILLUSTRATION 6.1 Model of marital satisfaction tested in questionnaire study.



words, a spouse's performance in his or her family roles probably influences both partners' sense of marital satisfaction, but the opposite is also probably true, that is, the degree of marital satisfaction probably influences how a partner performs his or her marital duties.

Identification of Population and Sample Selection

Decisions about what population to study usually are based on these functions: the topics of interest, the degree of generalization, the researchers' desire, and available resources. The population of interest may be a town, city, county, state, region, or nation, or some subsample or subgroup within these units, such as selected ethnic, religious, or age groups. Deviant populations, occupational groups, voluntary associations, and students are often the subjects of survey research.

In the example of the marital satisfaction study, the researchers wanted to generalize their findings to the American family but had only sufficient funds to study families in Utah. Thus, they defined married couples living in Utah as the population to be studied.

Generally, a target population is too large to survey in its entirety, and it is not necessary to study an entire population to describe it or to test hypotheses about it. Instead, a sample of the target population is selected to represent the entire group. Types of samples and how to draw them were discussed in Chapter 3.

For the study of marital satisfaction in Utah, the sampling frame was the twelve telephone books that covered the state. The number of names, excluding businesses, contained in the twelve books was calculated and a systematic random sample selected. These procedures produced a random sample of 2,227 households.

Questionnaire Construction

Most questionnaires contain items designed to measure the dependent variable(s), independent variables, and a fairly standard set of **demographic items**. Much of the discussion in the preceding chapter on interview schedule construction also applies to questionnaire construction. In addition, there are several principles that apply specifically to the construction of questionnaires.

First, the items in the questionnaire must be stated simply. They cannot be as complex as items in a face-to-face interview for there is no one present to offer definitions, clarifications, or probes. On the other hand, they need not be as simple and readily understood as items in telephone interviews.

Second, it is usually best to minimize the number of open-ended questions in a questionnaire. Most people express themselves better in speaking than in writing, and consequently open-ended items in ques-

tionnaires yield incomplete, inconsistent answers. Illegible handwriting and poor spelling contribute to problems in coding and interpreting responses to open-ended questions. Many survey researchers include a few open-ended items to provide illustrative detail or to highlight the range of interpretations about variables central to the study, but they also include measurements of the same substantive topics framed in structured, closed-response items.

Screening questions in an interview or questionnaire identify certain subgroups to whom special questions not applicable to all respondents are addressed. For example, a question on marital status may be used to identify married people who then are asked a series of questions about their spouses, children, or marriages. Unmarried respondents are instructed to skip the section of the questionnaire aimed at the presently married.

Screening questions are useful but should be used sparingly, as they can be confusing. Each "screen" requires an explicit set of instructions and the set of questions that apply to each kind of respondent must be clearly marked. The use of a box to highlight a set of questions following a screen is demonstrated in Illustration 6.2. Question 17 asked about marital status. If the respondent was married or remarried, he or she was requested to respond to questions 17a–d. Question 17d is another screen, identifying respondents who had an employed spouse, and those who did were to answer questions 17e–g.

Some questions, such as reports of past behavior, require a specific time frame if they are to provide useful data. For most purposes an item such as "How often have you smoked marijuana?" is too vague. Lacking a specific time frame in the question, people will answer in many different time frames, thus creating data that are, at best, inconsistent and, at worst, too ambiguous to be aggregated in analysis. Possible appropriate time frames might be "ever," "in the past five years," "in the past year," "in the past month," and so on. To measure the frequency of behaviors that happen often requires a short time frame, partly because respondents can't remember or estimate high frequencies over long periods very accurately. On the other hand, behaviors that occur infrequently generally require a longer time frame. Thus, a question on how often a college student has studied for an exam would include a shorter time frame than would an item on how often, if ever, he or she has committed robbery.

Finally, if there are sensitive questions, an appropriate context must be created for them. A question without a proper introduction may be objectionable to some respondents, and in a questionnaire survey there is no interviewer available to smooth ruffled feathers. The researcher must anticipate as many problems as possible and defuse them in advance. Dillman (1978: 106) gives an excellent example of this principle. Take the question,

ILLUSTRATION 6.2 Examples of screening questions.

17. WHAT IS YOUR PRESENT MARITAL STATUS?

1. Single, never married
2. Married
3. Remarried following divorce
4. Remarried following widowhood
5. Divorced
6. Widowed
7. Permanently separated
8. Living together

IF MARRIED OR REMARRIED:

17a. HOW FAR DID YOUR HUSBAND/WIFE GO IN SCHOOL?

1. 0-6 years
2. 7-9 years
3. Some high school but did not graduate
4. High school graduate
5. Some college but did not graduate
6. College graduate
7. Some graduate work but did not graduate
8. Completed graduate work

17b. EVERYTHING CONSIDERED, HOW HAPPY IS YOUR MARRIAGE?

1. Very happy
2. Happy
3. So-so
4. Unhappy
5. Very unhappy

17c. DO YOU HAVE ANY CHILDREN?

1. No
2. Yes

IF YES, WHAT ARE THEIR AGES? _____

17d. DOES YOUR HUSBAND/WIFE WORK?

1. No
2. Yes

IF YES:

17e. WHAT KIND OF WORK DOES YOUR HUSBAND/WIFE DO?

17f. WHAT KIND OF BUSINESS OR INDUSTRY IS THIS?

17g. WHAT WAS YOUR HUSBAND'S/WIFE'S INCOME LAST YEAR FROM HIS/HER EMPLOYMENT (BEFORE TAXES)?

- | | |
|------------------------|--------------------------|
| 1. Under \$3,000 | 7. \$13,000 to \$14,999 |
| 2. \$3,000 to \$4,999 | 8. \$15,000 to \$16,999 |
| 3. \$5,000 to \$6,999 | 9. \$17,000 to \$18,999 |
| 4. \$7,000 to \$8,999 | 10. \$19,000 to \$24,999 |
| 5. \$9,000 to \$10,999 | 11. \$25,000 to \$50,000 |
| 6. \$11,000 to 12,999 | 12. Over \$50,000 |

The church is a parasite on society. Do you agree or disagree with this statement? 1. Agree 2. Disagree

If this question were asked without any background, some people might object, refuse to complete the questionnaire, or even write nasty comments in the margin. However, placing it in an appropriate context can make the question acceptable to almost everyone. An explanation of why the question is being asked, and the inclusion of a positive item about organized religion, will assure most respondents that the researcher is interested in an objective assessment of attitudes about churches and is not anti-religious. Such an explanation might read:

Next I want to ask how you feel about the relationship between the church and society. Here are various opinions, both negative and positive, that we have heard people give on the topic and we would like to know whether you agree or disagree with each.

The church teaches people to help one another. Do you agree or disagree with this statement? 1. Agree 2. Disagree

The church is a parasite on society. Do you agree or disagree with this statement? 1. Agree 2. Disagree

Almost every questionnaire, regardless of the topic being studied, must include a set of demographic or background questions. These questions focus on personal characteristics that are widely accepted as relevant to most research objectives. They include standard items on gender, age, marital status, ethnicity or race, education, employment, occupation, income, and sometimes religion, housing tenure (rent or own), and household composition (who lives there in addition to the respondent). The Social Science Research Council (1975) recommends not only a common set of background characteristics, but also the use of a standard format so that accurate comparisons can be made between studies. This is not always possible, because some researchers will need greater detail on some demographic items than the standard form provides. For example, the level of detail on ethnic origin generally includes only the six major categories shown in the Brief Version section of Illustration 6.3. However, a researcher specializing in the influence of ethnicity may require much more detail, perhaps something on the order of that shown in the 22 categories of ethnic origin in the Detailed Version section of Illustration 6.3. Often such a detailed list can be made to fit the standard categories if detailed responses are chosen that can be combined into the more general categories. In sum, it may be necessary to ask some demographic items in a unique way, but the researcher should use the suggested list of characteristics and standard format when possible.

The order in which questions are presented in the questionnaire seems to make a difference, but there is disagreement about which questions should be placed where, and why (Sudman and Bradburn, 1982). Some prefer to place the demographic items early in the questionnaire and more

ILLUSTRATION 6.3. Two ways to ask ethnic origin: brief version and detailed version.

Brief Version

TO WHICH GROUP DO YOU BELONG?

1. White
2. Black
3. Asian American
4. American Indian
5. Mexican American
6. Other

Detailed Version

WHAT IS YOUR ORIGIN OR DESCENT?

- | | |
|------------------------------|----------------------|
| 1. American Indian | 12. Korean |
| 2. Central or South American | 13. Mexican |
| 3. Chicano | 14. Mexican American |
| 4. Chinese | 15. Negro or black |
| 5. Cuban | 16. Polish |
| 6. English | 17. Puerto Rican |
| 7. French | 18. Russian |
| 8. German | 19. Other Asian |
| 9. Irish | 20. Other Spanish |
| 10. Italian | 21. Other |
| 11. Japanese | 22. Don't know |

sensitive items toward the end, whereas others prefer the opposite. Lacking hard and fast rules, the researcher must make descisions about item order taking into account the topic being studied and the nature of the population to whom the questionnaire is being sent. If the topic is one of high interest and not particularly sensitive, the best strategy may be to start with questions on the main topic, thus assuring the highest completion rates for these early items. On the other hand, if the topic is not particularly interesting or if it is a sensitive subject, then generally a better strategy would be to start with the demographic questions to get the respondent involved in the questionnaire and then move to the more sensitive questions.

The first page of the questionnaire should tell the respondent how to fill it out. A sample cover page from a questionnaire survey of religious beliefs and practices in Middletown during the spring of 1978 is presented in Illustration 6.4. The instructions explain that the questionnaire contains three sections, and the nature of each section is described. All potential respondents were asked to complete sections I and II, and church members were asked to complete section III as well. Respondents were also told how to answer, by circling or checking the best single answer to each

ILLUSTRATION 6.4 Questionnaire cover page explaining study and instructing how to complete questionnaire.

RELIGION IN MIDDLETOWN

This questionnaire is designed to determine how people in Middletown feel about religion. It contains three sections: Section I asks social background information which permits comparative analysis. For example, the responses of those who are married can be compared to those who are single, those with children to those without children, etc. Section II concerns general attitudes towards various types of religion and religious practices. These attitudes are quite general and everyone is asked to respond to this section also. The final section (III) is included for those people who identify with a specific denomination and focuses on specific religious activities. Would you please respond to the first two sections and to Section III if you identify with a specific denomination.

Most questions can be answered by circling or checking one answer, so please don't mark more than one answer unless the instructions ask for more than one. If you wish to explain an answer or offer a comment, use the margin or attach an additional sheet. Remember that your answers are confidential and will be used only as data in the research report. Thank you.

question. They were also encouraged to write comments in the margins of the questionnaire or on a separate sheet of paper attached to the questionnaire.

Drafts of questionnaires should be pretested, often many times, before being printed in final form. Dillman (1978: 156) suggests that the questionnaire be pretested in three separate populations. First, a number of people drawn from the population to be surveyed (not part of the sample) should be asked to complete the questionnaire. Second, professional social scientists, trained and experienced in questionnaire construction should be asked to complete and evaluate the questionnaire. Finally, a group of potential users of the study, such as politicians, government officials, or agency administrators, should review the instrument, thus increasing the odds that the findings will be used in some way. The pretesting process identifies ambiguous questions or misleading instructions and sometimes suggests omitted items that should be included as well as included items that should be deleted. Also, the apparent effects of the way the questions are ordered can be determined and sections or items rearranged. It is essential that the people responsible for the construction of the instrument have some direct experience in the pretest, both by completing the instru-

ment themselves and by carefully debriefing the people who completed the questionnaire as part of the pretest.

The marital satisfaction study described above followed these procedures and the outcome was a 14-page questionnaire containing 57 questions. Several of the questions asked for information about seven different family roles, so that the questionnaire approached the maximum acceptable length, despite the modest number of questions. The only open-ended questions were two that asked how respondents felt about their marriages. The introductory comments on the cover page stressed the importance most people attach to family life and how vital it was to know how families were changing. Respondents were asked to answer every question, to give more than one answer only when requested to do so, and, if they had comments, to write them in the margins. Finally, they were assured that their answers were strictly confidential.

The questionnaire was reviewed by four sociologists at two different universities and pretested on 15 families. As a result of these pretests, there were some minor modifications of the questionnaire.

Administering the Questionnaire

When the instrument is completed, the first step in data collection is deciding how to distribute and collect the questionnaire. There are two popular strategies. First, the questionnaire may be hand-delivered to the individual respondents. In the project discussed earlier testing the relationship between religiosity and delinquency, the questionnaires were delivered to the respondents' homes and later picked up by a research assistant. Approaching respondents in groups is another efficient way to deliver questionnaires. For example, in a national study of the quality of American education (Coleman et al., 1966), questionnaires were administered to students in selected classrooms.

The second strategy, and probably the most popular, is to mail the questionnaires. If the respondents can be located in groups or live close to each other, it is probably faster and cheaper to hand-deliver questionnaires, but if the respondents are scattered over a large geographical area, a mail survey is much more efficient.

The second step in data collection is to introduce the study to potential respondents. Announcements can be made in local newspapers or by radio alerting the public that the study is commencing. An announcement may be mailed to potential respondents telling them a little about the project and indicating that a questionnaire will soon be delivered to them. Finally, the researcher may introduce the study personally at the time the questionnaires are delivered. In most mail surveys the cover letter provides the introduction.

The introduction, whether recited by the researcher delivering the

questionnaire, printed on the front of the questionnaire, or contained in a cover letter mailed to the respondent, should first explain the general topic of the study. At times it may be necessary to identify the area of interest somewhat vaguely so as to reduce the likelihood of asocial desirability biases in either direction. For example, an introduction explicitly explaining that a study is testing the relationship between religion and delinquency might raise conflicts between the respondents' desire to answer accurately and their sense of loyalty to their church. Explaining the study in more general terms, such as explaining that it is a descriptive study of adolescence, may decrease the likelihood of social desirability biases. Less sensitive topics may be explained more specifically and openly. For instance, in a study of community attitudes about the major nuclear weapons assembly plant in the United States (the Pantex Plant in Amarillo, Texas), the researchers explained that the study was a part of an environmental impact assessment of a proposed expansion of this facility. Whatever the level of detail, it is critical that the introduction stress the significance of the study to science, to society, and, if possible, to the individual respondent.

One way to emphasize the importance of the study is to have a high-status sponsor. Being able to say that a project is affiliated with universities, state or federal agencies, or respectable voluntary associations increases its significance for many people. Both the religiosity-delinquency study and the Pantex Nuclear Assembly Plant study had high-status sponsors, the former the Family Research Institute at Brigham Young University and the latter the Los Alamos National Laboratory and the Department of Energy.

The individual respondent's importance to the study should also be clearly stated. This is generally accomplished by noting that the respondent is part of a scientifically selected sample which was chosen to represent a wide variety of opinions and characteristics.

The introduction should end in a request for participation. The respondent is asked if he or she is willing to complete the questionnaire. The cover letter sent to a sample of respondents in a mail survey on working women in Middletown is presented in Illustration 6.5 as an example of an introduction. The study was loosely described as a "survey of women in Middletown" and there was an appeal to the importance of updating the classic 1924 study of Middletown. The identification number stamped on the questionnaire was explained as necessary to keeping track of whether respondents had completed and returned the questionnaires, and the confidentiality of all data collected was stressed. The letter was personally signed by one of the senior staff members.

Another strategy to motivate respondents to participate is to offer a modest incentive. Some researchers provide a ball point pen, a key chain, book mark, or some cash as a "token of appreciation" for filling out the questionnaire. Even if most people really don't want the proffered trinket

ILLUSTRATION 6.5 Cover letter used in mail questionnaire study of women in Middletown.

MIDDLETOWN III PROJECT
113 N. Council Street
Muncie, Indiana 47305
Telephone (317)289-6072

Dear

We are writing to ask your assistance in a study of women in Middletown. As part of the study of life in Muncie or Middletown, we are conducting a survey of the experiences of women, including those in the labor force. The enclosed questionnaire has been carefully designed to provide valuable scientific information about the experience of women and at the same time, we hope, to be meaningful and interesting.

The survey is being conducted by the Center for Program Effectiveness Studies, a research institute of the University of Virginia, as part of a four year study of Middletown, U.S.A. As you probably know, Muncie was the site of the original Middletown Study in the early 1920's. This study hopes to be able to document how life in Middletown has changed since the original Middletown Study in the early 1920's.

Your name was selected at random from all women in Muncie. Because you are one of a scientifically selected sample, your response is very important. We are interested only in combined statistics and ask that you not put your name on the questionnaire. To help us keep track of the questionnaires as they are returned, we have instead stamped numbers on the cover page.

If you would rather not participate in the study would you please return the questionnaire with the notation that you do not want to be a part of the study. We can then remove your name from the sample.

Only through studies such as this one can we know with any degree of certainty how women feel about life in Middle America. We strongly urge that you complete the enclosed questionnaire at your earliest convenience and return it in the postpaid envelope provided. If you have any questions, please call me at 289-6072. Thank you very much for your cooperation in this study.

Best regards,

Bruce A. Chadwick
Project Director, Middletown Study

or coin, research has shown that giving a gift creates a feeling of obligation and raises completion rates. Another strategy is to promise an incentive or reward when the respondent returns the completed questionnaire. The authors conducted a study of the relationship between education, including placement in foster homes, and the lifestyle of Indian Americans. To obtain data from respondents who had lived off-reservation and beyond the reasonable range of an interviewer, the authors offered a substantial monetary incentive (\$25). Occasionally a publishing company will offer faculty members a free book if they will respond to a questionnaire about what they

prefer in texts for a specific course or subject. Coupons for a free soft drink, hamburger, movie ticket, or the like can be used to increase the response rate. A popular incentive is to promise respondents a copy of the results of the study.

In a questionnaire study there is a limited contact between the researcher and the respondents and so there must be follow-up contacts to encourage people to complete the questionnaire. An initial mailing will usually produce completed questionnaires from under 20 percent of a sample. Of course, there is a much higher initial response rate if a research assistant hand-delivers a questionnaire and waits while the respondent completes it, but such questionnaires are necessarily quite brief. When questionnaires are administered to members of some organization or to students during a regularly scheduled meeting time, a very high response rate is achieved. If group administration is not possible, personal visits, telephone calls, or multiple follow-ups by mail are necessary to improve the completion rate. Follow-up programs must make additional copies of the questionnaire available to replace discarded or lost ones. Decisions about the kind and number of follow-ups in questionnaire studies are based on projected costs of additional personal visits, the availability of telephone numbers, current addresses of respondents, the response rate already achieved, and obtrusiveness (i.e., nuisance level as perceived by potential respondents already contacted) of additional follow-up efforts.

Mail questionnaire surveys are popular because they are efficient and relatively inexpensive, especially when respondents are widely scattered geographically or when large samples are needed. Also, most people have a mailing address and can be contacted via the mail. Persons who work nights, travel a great deal, or otherwise are unavailable to interviewers or telephone surveys can all be reached by a mail survey.

To conduct a mail survey, one needs the mailing address of potential respondents. City directories, telephone books, student directories, or membership lists of organization are frequently used as sampling frames and sources of addresses.

The initial mailing generally includes a **cover letter** (see Illustration 6.5), a copy of the questionnaire, and a self-addressed, stamped return envelope. The cover letter should be personalized as much as possible. The letter should open with the respondent's name and should be individually signed by the researcher with a pen (or color of ink) that makes the personal signature obvious. An incentive may be offered in the cover letter or included in the packet, if the researcher desires. If information is desired from more than one member of the household, such as both husband and wife, or parents and children, a questionnaire and a return envelope for each should be included. The return envelope usually has a printed business reply "postage paid by addressee" label so that postage

fees are paid only for the envelopes returned. This procedure is far less costly than affixing a stamp to the return envelope, thus paying the price of the stamp whether the return envelope is used or not. The packet should be sent first-class mail so that it will be forwarded to respondents who have moved or returned to the researcher if undeliverable. Also, first-class mailing, especially if the outer envelope is not the typical brown or manila "bulk mail" type, tends to raise the status of the packet out of the "junk mail" category.

Approximately two weeks after the initial mailing, a postcard follow-up should be sent. Dillman (1978) recommends that the postcard be sent exactly one week after the first packet is sent, but we have found that waiting another week frequently significantly reduces the number of postcards that have to be sent. The postcard is a brief, modified version of the cover letter and encourages the respondent to fill out the questionnaire. The postcard contains the researcher's phone number and invites the respondents to call if they have questions or if they have misplaced the questionnaire. If the sample includes respondents outside the local telephone service, a toll-free number should be given.

When the response to the reminder postcard tapers off, usually in about two weeks, an entire new packet is sent, including a different cover letter, another copy of the questionnaire, and another return envelope. This second follow-up may be done three or four weeks after the initial mailing. The new cover letter acknowledges the earlier contact and states that the researcher is bothering the respondent again only because of the importance of the research project. Where appropriate, the second letter includes a strong appeal to respondents to do their part to contribute to science, to society, or to humanity. Some researchers send a second reminder postcard two weeks after the mailing of the second complete packet; others wait longer for responses to the second complete mailing and then send a third and final complete mailing. This final follow-up is sent when responses to previous mailings have stopped or slowed to a trickle. Dillman (1978) procedures specify that the third and final follow-up should be sent seven weeks after the first mailing. By waiting until more responses to previous mailings have appeared, the researcher reduces the size of the final follow-up mailing costs and the possibility of offending respondents who are slow to reply. The final mailing includes a new cover letter, questionnaires, and return envelope, sent by registered or certified mail. The certified/registered mailing stresses the importance of the study to the respondent. The costs of both registered and certified have risen substantially in recent years, but the increase in response rates usually justifies the expense. The new cover letter announces that with reluctance the researcher is asking one last time for the respondent's help, apologizes for the possible intrusiveness of the multiple mailings, but emphasizes that

the study justifies the effort. An alternative to this final mailing is a telephone call or visit to the respondent's home in which a polite personal final appeal is made.

In the marital satisfaction study, each household in the sample received an initial mailing containing copies of the questionnaire, two return envelopes, and a cover letter. The cover letter requested that the husband and wife independently fill out the questionnaire. If only one adult lived in the household, he or she was instructed to discard the extra questionnaire. Two weeks later a postcard follow-up urging participation was mailed first-class to those households that had not returned at least one questionnaire.

Six weeks after the initial mailing, a complete new packet, including a revised cover letter, two questionnaires, and two return envelopes, was sent to households from which there had been no reply. Ten weeks after the first mailing, another full packet was sent by certified mail to the households remaining on the "no response" list. Another six weeks were allowed after the certified mailing for late returns. The entire data collection took five months, from March 1 to July 30, 1974. The rate of response will be discussed in the next section.

Analysis and Report

The **response rate** is calculated by tallying completed questionnaires and dividing this figure by the number of potential respondents. Respondents unavailable because of death, serious illness, or moving without a forwarding address are deleted from the potential sample, producing a corrected potential sample that is used to compute the response rate. Dillman (1978) reviewed 48 mail surveys that had used techniques similar to those described above and found that response rates varied from a high of 95 percent to a low of 50 percent. The average was 74 percent.

To continue with the example of the marital satisfaction study, of the 2,227 households (families) selected, 148 had moved leaving no forwarding address. Twenty-five questionnaires were returned with the notation that the addressee had died. These 173 ($148 + 25$) households were removed from the sample, leaving 2,054 households. Completed questionnaires were obtained from 1,199 households, for a response rate of 58 percent. Two hundred and eight (10 percent) indicated their refusal to participate by calling or sending a note to the investigators. Many potential respondents (652, or 32 percent) did not respond to the initial contact or the three follow-ups and were counted as refusals. Matched questionnaires, that is, completed questionnaires from both the husband and wife, were obtained from 775 couples, and it was these 775 pairs of questionnaires that were analyzed to test the model of marital satisfaction.

The responses in the mail questionnaires are assigned a numerical

value or "coded," thus preparing the data for computer analysis. The data are then entered on computer tape and a data file constructed. The process of coding, entering the data into the computer, "cleaning" the data, and preparing the data file for analysis is discussed in detail in Chapter 13.

The final report attempts to resolve the questions raised in the statement of the problem. If the study was conducted to describe something, the report contains that description. If the study tested a possible relationship, the statistical results are interpreted and the level of support for the relationship is discussed. Also, a research report should describe new insights, new problems identified, unanticipated methodological problems or other ideas stemming from the research which may interest other scientists, and suggestions or recommendations about how the results might be useful to scientists, program administration, or people generally.

The marital satisfaction study was reported in the *Journal of Marriage and the Family* (Chadwick, Albrecht, and Kunz, 1976). The article explained that a couple's disagreement about marital roles (who should do what, who does what, and how well each does it) was directly related to low marital satisfaction. On the other hand, conformity by one spouse to the other spouse's expectations about family role performance was the strongest correlate of marital satisfaction.

V. AN ILLUSTRATIVE EXAMPLE: EQUALITY OF EDUCATIONAL OPPORTUNITY

The Civil Rights Act of 1964 instructed the U.S. Commissioner of Education to conduct a survey and report to the President and the Congress about the educational opportunities available to American students of different races, colors, religions, and national origins. The study (Coleman et al., 1966) compared the educational opportunities of six racial and ethnic groups: black Americans, Indian Americans, Oriental Americans, Mexican Americans, Puerto Ricans living in the continental United States, and white Americans. Four major questions were addressed:

1. To what extent were racial and ethnic groups segregated in the public schools?
2. To what extent did the schools offer equal educational opportunities?
3. What level of academic achievement was attained in the public schools?
4. What relationship was there between school characteristics (answers to questions 1 and 2) and academic achievement (answer to question 3)? (See 6.6.)

By congressional mandate, the population to be studied included all public schools (elementary, middle, junior high, and senior high schools) in the United States. A three-step stratified cluster sample was used to identify 4,000 schools, 67,000 teachers, and 900,000 students. Counties,

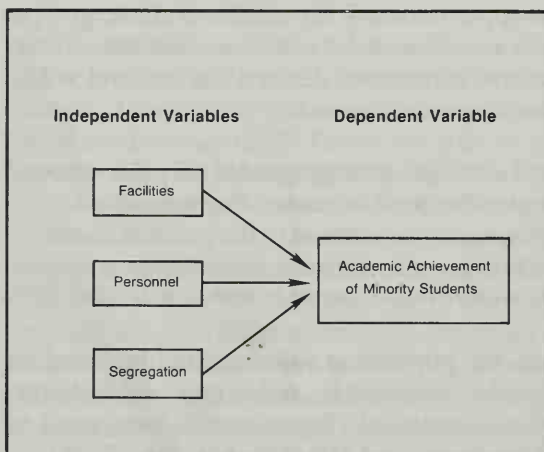
rather than school districts, were used as the first clusters of students to be sampled because the demographic information contained in the 1960 census allowed the stratification of the counties by population, region, and proportion of nonwhite population. The list of 3,100 counties in the United States was used as a sampling frame and a random stratified sample of 332 counties was selected.

The second step in the sampling process was to select a sample of schools within the previously selected counties. Rather than select a sample of all the schools in the selected counties, the researchers opted instead to draw a sample of high schools and then included all the elementary, middle schools, and junior high schools that feed students into the chosen high schools. A sample of 1,170 high schools was selected from the 4,522 in the counties.

The third step in the sampling process was to draw a sample of grades. Grades were not randomly selected, but rather the first, third, sixth, ninth, and twelfth grades were chosen, as they represent transitional points in the students' progress through the public school system. The sample, then, contained students in these five grades in the selected schools in the appropriate counties.

The U.S. Commissioner of Education sent a letter to the Chief School Officer in each state asking for their cooperation in negotiating with school districts and individual schools. When approval was obtained, the superintendent of the school districts and the principals of the selected schools were contacted. Figures on the number of students in the selected grades at each school were obtained and the appropriate number of questionnaires were mailed to the schools. September 28 and 30, 1965, were designated

ILLUSTRATION 6.6 Hypothesized relationships between school characteristics and academic achievement.



as the days when the questionnaires were to be administered. However, because of delays in shipping, not all schools were able to comply, and a few administered the questionnaires during the first week in October. The principal and teacher questionnaires were self-administered; student questionnaires and achievement tests were administered in class by teachers. A manual for each grade level accompanied the questionnaires and gave detailed instructions to the teachers on how the survey was to be administered.

These procedures produced a 67 percent response rate for schools. The researchers did not explain why one-third of the schools did not have their students participate in the study. Nevertheless, data relevant to the mandated issues were obtained from 640,000 students, 67,000 teachers, and 4,100 principals.

The amount of information obtained in this study via questionnaires was astonishing. The variables included the condition of classrooms, number of pupils per classroom, kind and quantity of school equipment, books, libraries, and other auxiliary facilities, teacher training, teacher experiences, teacher salaries, guidance and counseling programs, health programs, school lunch, curricula, school organization and administration, art programs, athletic programs, remedial programs, expenditures per pupil, nonenrollment, dropout rates, repeating of grades, overagedness, racial composition of classrooms, and many more characteristics. Appropriate academic tests were developed in specific subject areas for each of the grades. We can discuss only a few of the questions here. The interested reader is referred to the final report, *Equality of Educational Opportunity* (Coleman et al., 1966) for copies of the questionnaires and achievement tests used.

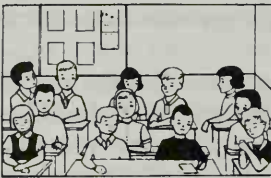
School principals were asked about the racial-ethnic backgrounds of the students they served. Teachers and older students were also asked how many students from the different racial groups were in their various classes. The younger students, grades 1 and 3, were shown the pictures in Illustration 6.7, portraying classes with varying proportions of dark versus light-skinned students and were asked to select the picture most like their current class, most like their class the previous year, and most like their group of "good friends." The information from all these sources was combined to estimate the degree of segregation in each school.

Principals, teachers, and students were asked to describe their schools. There were questions on many aspects of the curricula, facilities, programs, quality of teachers, and backgrounds of students. A sample of the information about school characteristics obtained is given in Illustration 6.8. This kind of information was used to assess the quality of educational opportunity.

Achievement tests in verbal ability, nonverbal ability, reading comprehension, mathematics, practical arts, natural sciences, social sciences, and humanities were prepared for the appropriate grades. First graders

ILLUSTRATION 6.7 Questions used to measure classroom segregation of first- and thirdgrade students.

Look around your class and then look at each of the pictures below. There are five questions about these pictures. For each question fill in the circle that has the same letter as the picture you choose.

A


B


C


D


E


39. Find the picture that looks **most** like the children in your class **now**.
 A ☐ B ☐ C ☐ D ☐ E ☐

40. Find the picture that looks **most** like the children in your class **last year**.
 A ☐ B ☐ C ☐ D ☐ E ☐

43. Find the picture which looks **most** like your good friends.
 A ☐ B ☐ C ☐ D ☐ E ☐

From Coleman, et al. *Equality of Educational Opportunity*, 1966, p. 618. Washington, D.C. U.S. Government Printing Office.

were tested in all the areas. An example of a verbal test and a nonverbal test for first-grade students is presented in Illustration 6.9.

The questionnaire and achievement tests were either answered on a machine-readable answer sheet or were themselves machine-readable. The information they contained was entered on computer files, data were cleaned (compared for illegitimate or inconsistent answers), and a complete data file was produced. Cross-tabulations between race, programs, and academic achievement were made, and summary tables and figures were created to portray the findings in an easily understood manner. A statistical technique known as **regression analysis** was used to measure the strengths of the relationships between segregation; characteristics of facilities, teachers, curricula, and programs; and academic achievement. Extensive **multiple regression analyses**, in which several independent variables are combined to try to account for changes in the dependent variable, were

ILLUSTRATION 6.8 Percent of pupils in secondary schools having selected characteristics, 1965.

CHARACTERISTIC	MEXICAN AMERICAN	PUERTO RICAN	INDIAN AMERICAN	ORIENTAL AMERICAN	BLACK AMERICAN	WHITE AMERICAN
Age of main building:						
Less than 20 years	48	40	49	41	60	53
20 to 40 years	40	31	35	32	26	29
At least 40 years	11	28	15	26	12	18
Average pupils per room	32	33	29	32	34	31
Auditorium	57	68	49	66	49	46
Cafeteria	72	80	74	81	72	65
Gymnasium	78	88	70	83	64	74
Shop with power tools	96	88	96	98	89	96
Biology laboratory	95	84	96	96	93	94
Chemistry laboratory	96	94	99	99	94	98
Physics laboratory	90	83	90	97	80	94
Language laboratory	57	45	58	75	49	56
Infirmary	65	77	77	69	70	75
Full-time librarian	84	93	85	98	87	83
Free textbooks	74	79	78	88	70	62
Sufficient number of textbooks	92	89	90	96	85	95
Texts under 4 years old	58	68	65	55	61	62
Average library books per pupil	8.1	6.2	6.4	5.7	4.6	5.8
Free lunch program	66	80	63	75	74	62

From Coleman, et al. *Equality of Educational Opportunity*, 1966, p. 11. Washington D.C.: U.S. Government Printing Office.

also performed. The results provided a cumulative measure of how much variation in academic achievement was accounted for by certain sets of school, teacher, and program characteristics. The authors cautioned against interpreting results from their survey as evidence of a causal relationship (Coleman et al., 1966: 290)—for example, that segregation “caused” low achievement among minority students. The final report makes it clear that there was a strong statistical relationship between segregation and achievement, but the findings did not demonstrate the chain of factors or events connecting school segregation to low achievement.

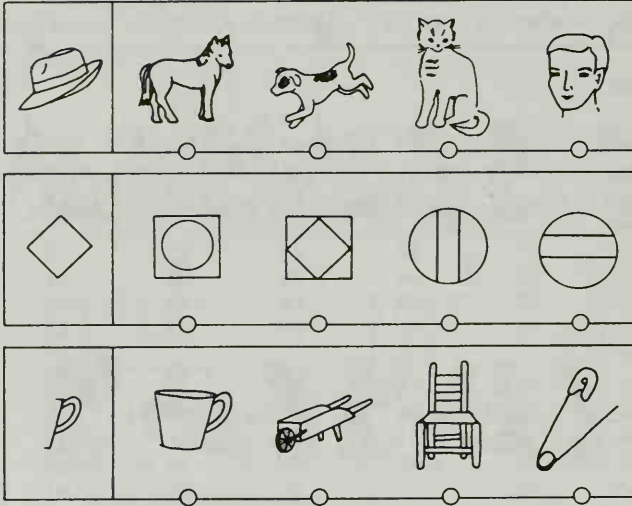
The Coleman report was prepared for Congress, but the authors realized that other government officials, as well as the general public, would be interested in the results. Therefore, they prepared a very readable report

ILLUSTRATION 6.9 Sample items from nonverbal and verbal achievement tests.

Part II of the test is a nonverbal classification test of 25 questions. Each question presents a stimulus picture and four response pictures. The

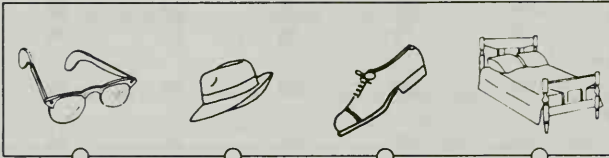
task is to select the one of the responses that "goes with the first picture." Four minutes are allowed, following practice questions.

Examples:

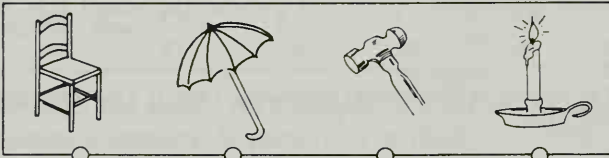


Examples:

Stimulus:
Hat



Stimulus:
Candle



Adapted from Coleman et al., *Equality of Educational Opportunity*, 1966, pp 578–79. Washington D.C.: U.S. Government Printing Office.

with numerous illustrations to present their findings graphically. The results have had a major impact on race relations in the United States and on American education.

According to the report, in 1965 most American children attended segregated schools, where almost all students were of the same racial background. For example, over 65 percent of black students in the first grade were attending schools where over 90 percent of the students were black. Although the percentage of black students in classes dropped a little

from the primary grades to junior high and high school, the majority of black students were shown to be in highly segregated schools.

The differences in school facilities were inconsistent: in some cases schools with high percentages of minority students were disadvantaged, in others they were favored. The authors concluded that for most minorities, especially blacks, the disadvantages outweighed the advantages and that the minority students in the United States attended inferior schools. They also concluded that minority students had less access to those curricular and extracurricular activities that were most related to high academic achievement. The same patterns of slight disadvantage for minority students appeared in the findings on quality of teachers; minority students had less able teachers than white students.

The data presented in Illustration 6.10 show the evidence documenting the lower academic achievement of minority students. When these test scores were translated into scores appropriate for each grade, it was discovered that the typical black high school senior was 4.1 years behind the white senior in verbal ability, 3.8 years behind in reading comprehension, and 5.0 years behind in math. In other words, the black senior had achieved only a seventh-grade level of math competence and an eighth-grade level of verbal and reading ability. The pattern was similar for other minority groups, with the exception of Oriental Americans, whose achievement in verbal ability and math was nearly equal that of their white peers, even though they were well behind in reading comprehension.

The regression and multiple regression analyses revealed that the

ILLUSTRATION 6.10 Academic achievement test scores obtained by students from different racial groups, 1965.

TEST	RACIAL OR ETHNIC GROUP					
	Puerto Ricans	Indian Americans	Mexican Americans	Oriental Americans	Negro	Majority
1st grade:						
Nonverbal	45.8	53.0	50.1	56.6	43.4	54.1
Verbal	44.9	47.8	46.5	51.6	45.4	53.2
12th grade:						
Nonverbal	43.3	47.1	45.0	51.6	40.9	52.0
Verbal	43.1	43.7	43.8	49.6	40.9	52.1
Reading	42.6	44.3	44.2	48.8	42.2	51.9
Mathematics	43.7	45.9	45.5	51.3	41.8	51.8
General information	41.7	44.7	43.3	49.0	40.6	52.2
Average of the 5 tests	43.1	45.1	44.4	50.1	41.1	52.0

factor most highly associated with academic achievement was the educational background and aspirations of fellow students. Those minority students who attended segregated schools were said to have a low level of academic achievement primarily because their fellow minority students were also from deprived backgrounds.

These results were interpreted by civil rights leaders, educators, government officials, the courts, and the public as powerful evidence that racial integration was the way to improve the academic achievement of minority students. Various strategies such as integrated magnet schools (a school with exceptional facilities and programs to attract students), feeder plans (students from white and minority elementary schools being "fed" into an integrated junior high school), and court-ordered busing programs were developed to integrate the public schools. Considerable integration has occurred since the Coleman report appeared, especially in the South. A dramatic increase in academic achievement of minority students was anticipated but has failed to materialize (Bahr, Chadwick, and Stauss, 1979). A possible reason that school integration has not increased the achievement of minority students very much is that the integration programs were based on the erroneous assumption that the results from the Coleman survey demonstrated a causal relationship between segregation and low academic achievement. The data did demonstrate that background of classmates was associated with academic achievement, but they did not demonstrate that integration would *cause* achievement to improve. It is also possible that there is a time-lag effect operating: that the positive results of school integration will not show up immediately, perhaps not for a generation or more. In any event, the Coleman study has had a significant impact on American education. Its findings justified a concerted effort to improve the quality of education of minority students, and the results of that effort will not all be apparent for many years.

VI. SUMMARY

A questionnaire study involves a respondent's filling out a self-administered "interview." The written instructions must explain how to complete the questionnaire and the questions themselves must be easily understood as there is no interviewer present to assist the respondent. Questionnaires can be delivered in a variety of ways, including hand-delivery to homes, hand-delivery to group gatherings like classrooms and work sites, and mailing to homes.

One major advantage of a questionnaire study is economy, as this type of research yields a maximum amount of data per research dollar. Another advantage is that the respondent can interrupt filling out the questionnaire to contemplate a question, to review records, or to consult

with another person. Some researchers argue that people will respond more truthfully to socially sensitive issues such as deviant behavior, sexual behavior, and extreme political views in a questionnaire than they will in an interview. On the other hand, there are researchers who contend that the opposite is true, that the interview is more sensitive in measuring embarrassing attitudes and behaviors. Both interviews and questionnaires are susceptible to inaccurate information and the researcher must guard against this.

The major limitations of a questionnaire are the number of questions that can be asked and the fact that no probing of insightful comments can occur. As a result, there is less depth in the information obtained compared with that obtained via interview. Another potential problem with a questionnaire survey is that persons other than the sampled respondents may complete the questionnaire. Nevertheless, if the researcher has a limited budget, needs data from a large sample, and is studying a research question with some theoretical and previous research support, a questionnaire study is appropriate. If the topic is a new area of study and if in-depth information and the pursuit of novel leads are important, then an interview study is called for.

Items contained in a questionnaire must be stated simply and only a few open-ended items should be included. Screening questions are used to identify respondents for which a section of the questionnaire is relevant so that other respondents can skip these. Questions should include a time frame. For example, rather than asking how often a respondent has smoked marijuana, the item should ask how often in the *past year* the respondent has smoked marijuana. Sensitive questions may need to be explained or placed in context with other items so that the respondents are not offended. A fairly standard set of demographic items is available and should be part of every questionnaire study. The order of questions has been hotly debated. The best strategy seems to be that if the study involves an interesting or nonthreatening topic, it should start with questions pertaining to it. On the other hand, if the focus is a boring or sensitive topic, then it probably is best to put the demographic items first. The cover page of the questionnaire should contain the specific instructions necessary to complete it.

Potential respondents can be made aware of the study via newspaper, radio, and television announcements or by a letter. The announcement should explain the purpose of the study as well as its contributions to science, to the community, and to the respondent. The questionnaires can be delivered by the researcher or assistants or by the mails. Whatever technique is used, follow-ups are necessary to increase the response rate. In-person call-backs, telephone calls, postcards, and letters can be used to encourage respondents to complete and return the questionnaire.

REFERENCES

- ALBRECHT, STAN L., BRUCE A. CHADWICK, and DAVID S. ALCORN, 1977. "Religiosity and Deviance: Application of an Attitude Behavior Contingent Consistency Model." *Journal for the Scientific Study of Religion* 16, No. 3: 263-274.
- BAHR, HOWARD M., BRUCE A. CHADWICK, and JOSEPH H. STAUSS, 1979. *American Ethnicity*. Lexington, Mass.: Heath.
- CHADWICK, BRUCE A., STAN L. ALBRECHT, and PHILLIP R. KUNZ, 1976. "Marital and Family Role Satisfaction." *Journal of Marriage and the Family* 38, No. 3: 431-442.
- COLEMAN, JAMES S., et al., 1966. *Equality of Educational Opportunity*. Washington, D.C.: U.S. Government Printing Office.
- DILLMAN, DON A., 1978. *Mail and Telephone Surveys: The Total Design Method*. New York: John Wiley.
- MANGIONE, THOMAS W., RALPH HINGSON, and JANE BARRETT, 1982. "Collecting Sensitive Data." *Sociological Methods and Research* 10, No. 3: 337-346.
- SOCIAL SCIENCE RESEARCH COUNCIL, 1975. *Basic Background Items for U.S. Household Surveys*. Washington, D.C.: Center for Coordination of Research on Social Indicators, Social Science Research Council.
- STROMMEN, MERTON, MILO BREKKE, RALPH UNDERWAGER, and ART JOHNSON, 1972. *A Study of Generations: Report of a Two-Year Study of 5,000 Lutherans Between the Ages of 15-65, Their Beliefs, Values, Attitudes, Behavior*. Minneapolis: Augsburg.
- SUDMAN, SEYMOUR, and NORMAN M. BRADBURN, 1982. *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco: Jossey-Bass.
- WISEMAN, FREDERICK, 1972. "Methodological Bias in Public Opinion Surveys." *Public Opinion Quarterly* (Spring): XXXVI, 105-108.

Experimental Research

seven

I. Introduction

II. Types of Experiments

Laboratory experiments

Field experiments

Natural experiments

III. Strengths and Weaknesses of Experimental Research

IV. Experimental Subjects

V. Steps of Experimental Re- search

Statement of research problem

*Development of operational
definitions*

type of experiment

experimental task

specific operational definitions

Selection of subjects

Experimental design

internal validity

design

Debriefing

Pilot study

Analysis and report

VI. Illustrative Experiments

*Milgram's obedience to au-
thority experiment*

statement of research problem

operational definitions

selection of subjects

experimental design

debriefing

pilot study

analysis and report

*Behavior modification of un-
derachievement and disrupt-
tive behavior experiment*

statement of research problem

operational definitions

selection of subjects

experimental design

debriefing

pilot study

analysis and report

*Guaranteed income experi-
ment*

statement of research problem

operation definitions

selection of subjects

experimental design

debriefing

pilot study

analysis and report

VII. Summary

I. INTRODUCTION

Experimental research involves selecting a dependent variable and one or more independent variables that are hypothesized to be causally related to it. The independent variable(s) are then manipulated by the experimenter under carefully controlled conditions to determine any changes in the dependent variable.

For example, Day (1971) hypothesized that a supervisor’s leadership style would influence worker productivity. He was also interested in workers’ feelings about different leadership styles (see Illustration 7.1). To test

ILLUSTRATION 7.1 Hypothesized relationship between supervision style and worker productivity and feelings toward supervisor.

<u>INDEPENDENT VARIABLES</u>	<u>DEPENDENT VARIABLES</u>
Leadership or Supervision Styles 1. Close supervision 2. Supportive supervision 3. Punitive supervision	Worker productivity Feelings toward the supervisor

his hypotheses, male undergraduate college students were recruited and paid to participate in an experiment. Upon arrival at the laboratory they were instructed (incorrectly) that the purpose of the experiment was to look at the leadership techniques of persons from different social backgrounds. They were asked to use chemistry demonstration kits with balls of different size and color to build rather complex molecular models. The balls, representing different atoms, had varying numbers of holes drilled in them into which two different lengths of pegs and springs could be inserted. A moderate incentive of 20 cents per model was offered to increase motivation.

Productivity was measured by counting the number of connections made between balls. The subjects were instructed that they would be working under the direction of a supervisor from a specific social class background. In addition to building the models, part of their assignment was to evaluate the supervisor's performance. They did this by pressing one of two "gas pedal" type switches under the work table with their feet. The subject could express negative feelings by pushing one switch and positive feelings with the other.

In reality the supervisor was a trained confederate who alternately played three leadership roles. At specified times he supported the efforts of the subjects by making 27 supportive comments such as "You have your materials well organized," or "You read blueprints really well." At selected times the confederate used a close leadership style, being very task oriented, by offering 80 comments such as: "If you insert the springs by twisting them counterclockwise, they go in easier," or "By holding the ball in your left hand and spring or peg in your right hand, you can build the models more quickly." The third leadership style was punitive; the supervisor verbally abused the subjects. He made 37 cutting remarks such as: "You act like you have two left hands," or "You are slower than my four-year-old son." The impacts of these three leadership styles acting singly and in certain combinations were determined by counting the number of connections made in the completed models and the number of times the positive and negative switches were pressed.

Day found that supportive leadership did not alter productivity, that close supervision reduced it, and that punitive leadership significantly increased it. However, before shop supervisors institute new get-tough personnel policies, they should be warned that the punitive style also generated strong aggressive feelings toward the supervisor.

This chapter will discuss the basic strategies for conducting an experiment. These include assessing whether experimentation is an appropriate means to test the research question, deciding what type of experiment to conduct, determining who and how many subjects are needed, designing the experimental setting, collecting the data, debriefing the subjects, analyzing the results, and writing the report.

II. TYPES OF EXPERIMENTS

Laboratory Experiments

In the laboratory, the researcher can control the immediate physical characteristics as well as the social environment by determining the number of subjects and confederates present, whether they can see or talk to each other, and what information is given to them. The strength of the laboratory experiment is this control which assures the researcher and the reader that any changes observed in the dependent variable derive from the independent variable(s). The weakness of such experiments is the difficulty of creating "reality," or a real-life situation in the laboratory. For example, intimate family relationships, long-term work situations, and other complex relationships are difficult to create artificially.

Field Experiments

Unlike the laboratory experiment, the **field experiment** is conducted in real social settings such as schools, factories, mental hospitals, prisons, churches, clubs, or homes. The strength of this type of research is that subjects are more likely to behave in a normal fashion in the real setting.

Also, field experiments generally last longer than laboratory ones.

ILLUSTRATION 7.2 This cartoon illustrates a laboratory supposedly providing the experimenter excellent control over the important independent variable.



Tests of teaching innovations, work procedures, or rehabilitation programs may take weeks, months, or even years.

A problem in field experiments is that the researcher gives up some control of the research context. Principals, teachers, custodians, supervisors, foremen, fellow students, co-workers, and others may introduce unexpected changes in the experimental setting which interfere with testing the hypothesized relationship.

Natural Experiments

Sometimes the researcher may wish to test the relationship between a dependent variable and an independent variable that he or she cannot control or manipulate. Some independent variables, like tornadoes and earthquakes, cannot be created with current technology and some cannot be manipulated because of moral and/or ethical considerations. Natural events such as floods or negative social events like closing a manufacturing plant cannot be inflicted on research subjects. The **natural experiment** involves preparing instruments to measure the dependent and independent variables, obtaining baseline levels of them, and then waiting for the independent variable to occur or change *naturally*.

In certain disaster research centers, projects are designed and personnel are prepared to begin field work at disaster sites where communities are suddenly faced with the consequences of a burst dam, tornado, or race riot. Occasionally researchers learn ahead of time that disruptive social events such as major layoffs at a large manufacturing plant or the closing of a military base are going to happen and they can collect baseline data about psychological, familial, and social relationships before the event. Later, following the disruptive "stimulus," observers can systematically assess the consequences of the event by collecting information on individuals, families, or the community as a whole and contrasting the findings with the baseline information.

A natural experiment is the only way to test experimentally the effects of some independent variables. On the other hand, the lack of control over the independent variable is a serious limitation of natural experiments. The experimenter must carefully prepare and then wait for the forces of nature or society to activate the independent variable. Often the unique characteristics of a disaster or social disruption are not very well suited to the needs of the experimenter.

III. STRENGTHS AND WEAKNESSES OF EXPERIMENTAL RESEARCH

As was noted earlier, the strength of experimental research is its ability to test cause-and-effect relationships. In the productivity study, the experimenter concluded that punitive leadership style produced both high pro-

ductivity and powerful hostile feelings among workers. The results of experimental research can be utilized in social engineering to alter society to improve the quality of life of its citizens.

Despite the utility of experimental research in revealing cause and effect, most social research is nonexperimental and does not reveal unambiguous causal relationships. Unfortunately, social scientists, government administrators and policy makers, practitioners, and the general public often assume that causal linkages have been demonstrated, and they may design programs around inadequate knowledge about what the consequence of certain actions or policies may be. Most of the logic underlying anti-poverty programs developed during the "War on Poverty" during the 1960s was grounded in nonexperimental studies. As a consequence, expectations were raised, but few anti-poverty programs produced the anticipated effects.

A specific example of mistakenly assuming a causal relationship from nonexperimental research was the famous study referred to in the previous chapter, *Inequality in Educational Opportunities in American Society* (Coleman et al., 1966). This survey discovered that minority students in segregated schools did not score well on tests of academic achievement. Government and school officials, judges and lay people mistakenly concluded that reducing segregation would improve minority students' academic achievement. Legislation was passed, administrative policies implemented, and court orders issued to force the integration of public schools. Later experimental research, primarily natural experiments, clearly demonstrated that integration had little, if any, impact on the academic performance of minority students. So conclusive is the experimental evidence on this point that Coleman himself (1975) conceded the error of assuming the relationship as causal. These disappointing attempts to increase the quality of specific aspects of social life point out the importance of experimental research.

The productivity experiment mentioned above also demonstrated the problem of **external validity**, or the degree to which the findings of an experiment can be generalized to other populations and other settings. The question is whether punitive, supportive, and close supervision would have the same effect on the productivity of noncollege students in non-laboratory situations. For example, would punitive supervision raise the work output of machinists in a steel mill, secretaries in a bank, children in a public school, or members of the local Chamber of Commerce? One important determinant of external validity is the type and number of subjects. Day (1971) had trouble finding individuals willing to come to the university laboratory to spend two hours. His solution was to recruit students in an undergraduate social science class and to pay them to participate. There is a danger in assuming that since leadership style in-

fluenced the productivity of a rather modest number of atypical undergraduates, it will therefore influence the productivity of adults, younger people, and those with less education operating within families, schools, factories, offices, the military, churches, and other organizations. In order to maximize external validity, the experimenter must do all he or she can to use subjects who represent the populations they wish to generalize the results to.

The characteristics of the subjects used in an experiment are not the only factors affecting external validity. The experimental setting, how close it represents reality, may influence the results to the point that they cannot

ILLUSTRATION 7.3 This cartoon illustrates the problem of generalizing an experimental effect from one group of subjects (chickens, college students, etc.) to another (mice, adults, children, etc.).



"It cures it in chickens;
it causes it in mice."

be generalized to other settings. There is the question of whether some complex social situations can ever adequately be recreated in a controlled laboratory setting.

External validity should be considered at each step of the experiment, and a realistic experimental setting should be created and appropriate subjects recruited. In spite of all that the experimenter does, however, external validity can never be proven or logically justified (Campbell and Stanley, 1966: 5).

To generalize from one population in a given setting to another in a different setting requires the acceptance of unproven assumptions. This fact does not excuse the researcher from worrying about external validity, but rather indicates the necessity of doing everything possible to maximize external validity at every stage of a project.

Another problem in Day's productivity experiment was that the experimenter lied to the subjects about the purpose of the study and about who the supervisor-confederate was. Day assumed that if he had told them the truth about the project's purpose, the subjects might not have reacted "realistically" or "normally" to the leadership styles.

This is a painful dilemma, as strict ethics requires that subjects be fully informed about what is going to happen to them; at the same time, however, if they are informed then they may not behave normally. Some researchers (Rubin, 1983) have assumed that, because of concern voiced in the late 1960s, deception in experimental research has declined. Gross and Fleming (1982) as reported in Rubin (1983) shattered this illusion as they found that the percentage of experiments reported in the *Journal of Personality and Social Psychology* using deception rose from 41 percent in 1959 to 66 percent in 1969 and declined very little to 62 percent in 1978–1979. Rubin (1983: 74–75) examined a recent issue of the same journal and concluded that "All the old tricks were there—the false reports to subjects about 'intelligence test' results, the confederates (or 'stooges') who pose as fellow subjects and then follow scripts, the phony equipment proudly described by the researchers as 'an impressive array of switches and electrical hardware.' "

Finally, Day ran the risk, albeit a small one, that a subject would be emotionally upset because of a supervisor's punitive comments and suffer loss of self-esteem, have a nervous breakdown, commit suicide, or have a heart attack. Such extreme outcomes were highly unlikely in this particular study, but psychological damage is always possible when an experimenter places subjects under stress.

In deciding whether and how to conduct an experiment, the researcher must weigh advantages against problems and potential dangers. If, on balance, the benefits outweigh the costs, then the researcher must decide whether a laboratory, field, or natural experiment offers the best balance between possible costs and benefits.

IV. EXPERIMENTAL SUBJECTS

The purpose of social science research is to discover principles that can be used to explain, predict, and even control social behavior. Because of its power to test causal relationships, experimental research is highly esteemed in the scientific community. However, the external validity of relationships demonstrated by experimental research has often been limited because the subjects were atypical. Also, the small number of subjects normally involved even further limits generalizing of the findings to larger populations. People who volunteer to serve as subjects may be atypical in many ways. Anxiety about being a "guinea pig" or unwillingness to spend time required makes it difficult for researchers to recruit "average" subjects.

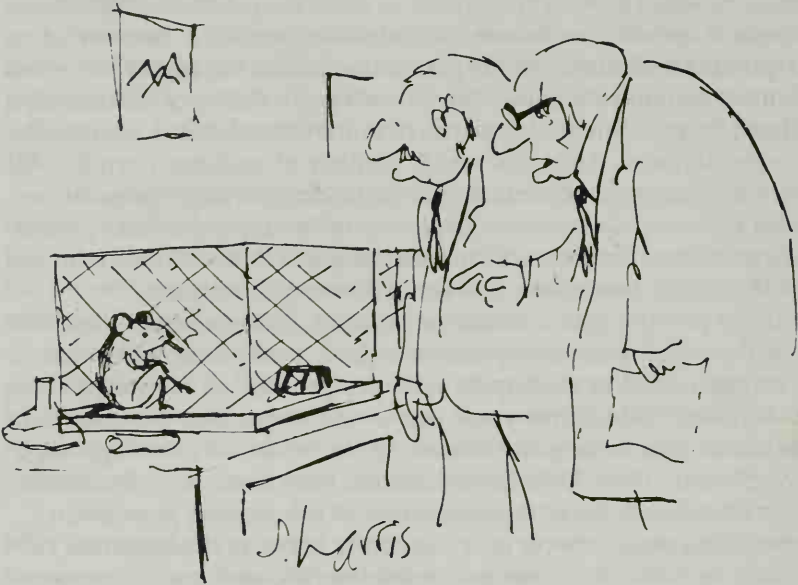
The usual practice is to use college students. College professors, who do most of the social science experiments, have ready access to students who can be persuaded to participate with the promise of extra credit or a better grade. More than thirty years ago it was noted that the science of human behavior was largely the science of the behavior of college sophomores (McNemar, 1946: 333). Experimenters have been strongly encouraged to recruit subjects more representative of the general population.

Students in public schools also frequently serve as experimental subjects. Inmates or patients in prisons, reformatories, and mental hospitals are occasionally utilized for they are captive groups to which powerful inducements may be offered. For example, prisoners willingly participate in rather dangerous medical experiments in hope that it will impress parole boards and hasten their release.

Rosenthal and Rosnow (1975) reviewed six studies that had examined psychology journals to identify the kinds of subjects used in experimental research in the 1960s. They learned that in 70 to 90 percent of the experiments reported in the journals the subjects had been college students. Little change had occurred in the recruitment of subjects since McNemar's warning three decades earlier.

The authors of this book reviewed experiments reported in three psychological and social psychological journals in 1981 to determine if the tendency to use college students as research subjects had changed during the 1970s. One hundred twenty-eight experiments were reviewed in the *Journal of Social Psychology*, *Journal of Experimental Psychology*, and *Journal of Research in Personality*. The results, presented in Illustration 7.5, are consistent with previous results; 70 percent of the experiments had used college students as subjects and an additional 15 percent had used public school students. Thus in 85 percent of the experiments reported in these three journals during 1981 the subjects had been students. A few studies (8 percent) had used adults recruited primarily from newspaper advertisements, and such volunteers do not represent the general population any better than do college or public school students. Other experiments had

ILLUSTRATION 7.4 This cartoon illustrates the experimenter's concern about the typicality of the subjects used in the experiment.



“He’s the typical American mouse—
likes a drink before dinner, smokes a little, watches TV . . .”

© Sidney Harris. From *All Ends Up*, Los Altos, Calif.: Kaufmann, Inc. Used with permission.

ILLUSTRATION 7.5 Research subjects in experiments reported in three social psychological journals,* 1981

SUBJECTS	PERCENT (N = 128)
College students	71
(social science students)	[39]
Public school students	16
Adults	8
Children	2
Employees	2
Alcoholics	2
	100

**Journal of Social Psychology*, *Journal of Experimental Psychology*, and *Journal of Research in Personality*.

recruited children from nonschool settings, studied employees in work situations, or experimented with clients of social service clinics.

The bias of generalizing to other populations from the results of experiments with college students varies with the topic being studied, but it is especially significant for socially learned behavior. For example, the reaction of college students to frustration or stress is probably quite different from the reactions of blue-collar workers, high school dropouts, successful businesspeople, middle-aged housewives, or senior citizens.

Research has discovered that college students differ in significant ways from their noncollege peers (Rubenstein, 1982). They have higher self-esteem, use drugs and alcohol less often, and are less likely to be married than other young people. In addition, young people have been found to be significantly different from older citizens. Rubenstein (1982) reported that youth and young adults, especially college students, are lonelier, often more bored, and unhappier with life than are older people. These differences led Rubenstein (1982: 83) to conclude that: "All of these demographic, intellectual, and psychological differences seem to make college students an atypical group of American adults."

A second important issue concerning subjects is the number used in a given experiment. Holmes (1979) reviewed the sample size of experiments reported from 1955 to 1977 in the four journals published by the American Psychological Association. The median number of subjects in individual treatment conditions was 14 and the median number for experiments was 48. Importantly, no increase in sample size was noted during the 22-year period. In a follow-up study ten non-American Psychological Association journals were analyzed for 1977 (Holmes, Holmes, and Fanning, 1981). The results were almost identical: a median of 15 subjects in individual treatment conditions and a median of 50 subjects per experiment.

The uncertainties and possible errors in generalizing the results from an experiment of 65 college students to people in the same town, state, or nation are readily apparent. Of course, the small "sample" of unique subjects does not accurately represent any wider population.

A common solution to the problem of obtaining research subjects has been to use volunteers. Many researchers have visited different groups and recruited volunteers. Others have advertised in newspapers or on radio for volunteers. Typically, in such advertisements the experiments are described in very vague terms and potential subjects may assume the experiment will be both enjoyable and educational. Such procedures have been somewhat successful in recruiting nonstudent subjects but have introduced a different bias. Rosenthal and Rosnow (1975) reviewed the literature comparing characteristics of volunteers and found that they are better educated, have higher social status, and are more sociable than nonvolunteers (Rosenthal and Rosnow, 1975: 195). Subsequent research has revealed even more differences between volunteers and nonvolunteers.

Rush, Phillips, and Panek (1978) discovered that, in addition to the personality differences noted by Rosenthal and Rosnow, volunteers were more adept than nonvolunteers at doing the research tasks. Apparently people who think that they can do well at the type of activity described in project advertisements are more likely to volunteer than those who feel they might not perform well.

The current emphasis on informed consent of subjects as an ethical issue was discussed in an earlier chapter. Here let us simply say that such requirements make it more difficult for researchers to select and recruit appropriate subjects. On the one hand, informed consent almost requires the use of volunteers, who have been shown time after time to be unrepresentative of the general population. On the other hand, it introduces possible biases in that subjects must be more aware and hence more biased about the study and may be more inclined to modify their behavior in line with their definition of the researchers' purposes.

The experimenter should make every effort, within budget constraints, to recruit representative subjects and to use as large a sample as possible. The consumer of experimental research should pay particular attention to characteristics and numbers of subjects.

V. STEPS OF EXPERIMENTAL RESEARCH

Statement of the Research Problem

As is true of all research projects, an experiment starts with a clear statement of what is to be studied. In the case of an experiment, not only is the dependent variable identified, but so is one or more independent variables. The reasons for hypothesizing a causal relationship between the independent and the dependent variables should be documented with reference to previous research. For example, a researcher interested in increasing the academic achievement of ninth-grade students may develop a hypothesis that matching the length of class periods to the attention span of a normal 14-year-old would raise the level of material comprehended. Previous research concerning the length of class periods with students in other grades and research about the attention span of ninth graders would have to be reviewed, and if the previous research supported the hypothesized relationship, the experimenter would proceed. If on the other hand, previous research refuted the hypothesis, the researcher probably would wish to modify the hypothesis in line with the findings of previous researchers.

Development of Operational Definitions

Type of experiment The researcher must decide what type of experiment he or she wishes to conduct. This decision is based on the risk to subjects, the need to control relevant situational variables, the ability to

create a realistic setting in the laboratory, access to a field setting, and the research budget. In the example of an experiment testing the relationship between length of class period and academic performance, we would opt for a field experiment because we want to test the hypothesis in a full academic year, and we will assume administrative approval for access to local junior high school students.

Experimental task The next step is to select an experimental task. Generally, some type of learning, problem-solving, or evaluation activity is used as the task. The experimenter's imagination is the only limit to kinds of tasks that might provide an adequate test of the hypothesized relationship. An example of an imaginative task is a fascinating experiment testing the relationship between TV violence and subsequent violent behavior. The task involved watching a preview of a segment of the TV show *General Hospital* (Milgram and Shortland, 1973). Later the subjects were requested to go to an office building to pick up a transistor radio promised them for their participation. A sign directed them to an office in which a situation was created allowing the subjects the opportunity to copy the violent behavior they had witnessed in the theater. In our present example, we would define normal ninth-grade learning activities in three academic classes—math, English, and science—as the experimental task.

Specific operational definitions Measurement of the dependent variable must be carefully planned. It may involve mechanical devices, such as event recorders, tape recorders, or video recorders; observation by carefully trained observers; or self-report by the subjects themselves. The productivity-leadership experiment (Day, 1971) operationally defined productivity as the number of connections made between the balls with pegs or springs. An accurate count was easy to accomplish. Feelings of approval or disapproval of the leader were operationally defined by the number of times the subjects pushed switches with their feet. The number of pushes was automatically recorded on an event recorder.

In the illustrative experiment, we would operationally define academic achievement as scores on the California Achievement Test (CAT), a standardized achievement test. A different version of the CAT would be given at certain points during the experiment.

Special attention must be given to the operational definition of the independent variable that is manipulated. First, an experiment needs significant variations in exposure to the independent variable. To experimentally expose the students to 50-minute, 53-minute, and 55-minute classes would not allow much of an impact to be made, as the lengths differ by only five minutes or less. At the same time, the experimenter needs to test increments of exposure that are practical. It is highly unlikely that any junior high will adopt five-minute or six-hour class periods. A reasonable

compromise would be to vary class length 20, 40, 60, 80, 100, and 120 minutes.

There is a temptation to shorten the experiment and reduce its cost by testing only the two extreme class times, 20 and 120 minutes. The problem with such an approach is that with only two or three different levels of the independent variable, it is difficult to ascertain its relationship to the dependent variable. In Illustration 7.6 it can be seen that if we had tested only the 20- and the 120-minute length, we would have wrongly concluded that length of class had no effect on academic achievement. By including the other units of time, we discovered that about 60 minutes is the length of class producing maximum academic performance.

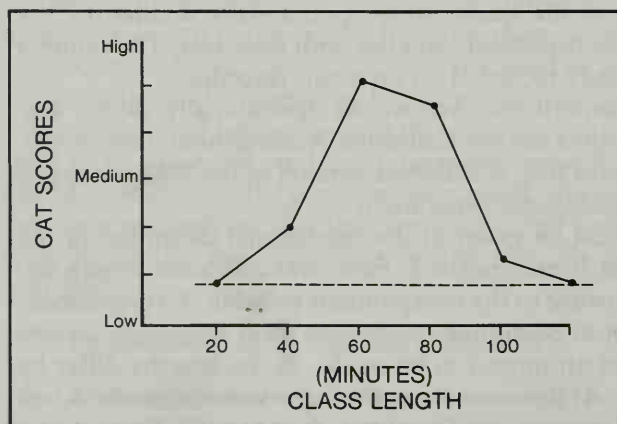
Selection of Students

The two essential questions in selecting subjects are what kinds of subjects and how many are needed. The answers to these questions are guided by the population the researcher wishes to generalize to, the nature of the dependent variable (i.e., is it influenced by social learning), and access to subjects.

In the length of class-academic achievement sample, we would use the entire ninth-grade class in the community's only junior high school. In a larger metropolitan school district with several junior highs, a sample of schools or classes would be more appropriate. The junior high to be used as the experimental site has a heterogeneous population with students from a wide range of social backgrounds, and the results would be generalized to similar school populations.

This hypothetical experiment would be sponsored by State Univer-

ILLUSTRATION 7.6 Hypothetical relationship between length of class and academic achievement.



sity, and its Institutional Review Board would have to approve the treatment of subjects. We would try to convince the IRB that "implied" **informed consent** would be sufficient. Implied informed consent would be obtained by sending a letter home to parents explaining the study and instructing them that if they have no objections to their child's participating, to throw away the letter. If they preferred that their child not participate, they should sign the attached form and return it to the school. This technique would obtain many more experimental subjects than would requiring the parents to return a signed permission form. Most parents ignore such letters, whatever their feelings about the experiment.

There are no hard and fast rules or formulas the researcher can use to determine who and how many subjects should be used. Generally, if the subjects are known to be representative of the population the researcher wishes to generalize about, fewer subjects are needed. If the hypothesized relationship between independent and dependent variables is based on a foundation of previous research, fewer subjects are necessary than if the research is exploring a new area. Also, if the initial findings of an experiment confirm the hypothesis, are consistent, and support previous research, fewer subjects are needed than if the initial findings are contradictory.

The selection of research subjects is a major problem in experimental research and no reasonable effort should be spared to obtain appropriate subjects. Ideal subjects are seldom available, and experimenters sometimes must make do with what the resources allow. On the other hand, difficulties in finding subjects should not serve as excuses to use subjects who are handy but inappropriate.

Experimental Design

Design refers to the overall strategy concerning the setting of the experiment, the assignment of subjects to experimental or to control groups, and the order of presentation of the independent variable. A good experimental design should maximize internal validity.

Internal validity Internal validity is the degree to which extraneous variables are controlled so that the researcher is confident that the changes in the dependent variable were caused by varying of the independent variable. This section draws heavily on the work of Campbell and Stanley (1966), who identified seven extraneous factors which, if not controlled, might produce change in the dependent variable unrelated to the effects of the independent variable. Each of these and one other extraneous factor will be discussed. We will also present several alternative experimental designs.

HISTORY. As used here, **history** refers to events or variables beyond the control of the investigator which influence change in the dependent

variable observations. In the example of the experiment on length of class periods, an announcement by the school district of a new competency examination required for graduation from junior high might produce a surge in academic performance which may mistakenly be attributed to duration of class period. Lengthy experiments requiring observations over several weeks or months are more vulnerable to the influence of historical changes than are brief experiments. Even in a one-hour laboratory experiment, however, a subject's laughing, fainting, or behaving in some unanticipated way can create an atypical history which influences the performance of other subjects on the dependent variable. In order to be significant, historical events must affect most of the subjects, as would be the case with the school district's new competency test. An event that affects only one or two subjects, such as parents offering their ninth-grade student a motorcycle if he gets better grades, usually will not bias the results and may be assumed to be cancelled out by actions of other parents reducing the achievement of their students.

MATURATION. *Maturation* refers to biological, psychological, and emotional processes that change subjects over time. In brief experiments, hunger, thirst, fatigue, or the need to visit a restroom may influence the subject's behavior. In longer experiments, like the nine-month class length-achievement example, maturation also refers to the physical and intellectual growth, improvement in coordination, and great emotional maturity that occur because the subjects age nine months during the course of testing.

TESTING. An initial observation may affect the dependent variable. In fact, this factor illustrates the Heisenberg principle that you can't study something without changing it. The *testing effect* is of special significance in experiments studying intelligence, achievement, and personality, because the process of taking an IQ, achievement, or personality test to measure student characteristics during the experiment gives the subjects practice in test taking and familiarizes them with the kinds of tasks or questions they may confront again at later stages of the experiment. Thus the subjects' scores on subsequent tests will be higher, but not because of the influence of the independent variable. In the earlier example of leadership and productivity, *practice* probably increased the rate of construction of the models independent of the style of leadership. In the length of class period-academic achievement experiment, giving the students several California Achievement Tests, even though several different versions of the test would be used, increases the risk that testing will affect the students' performance independent of class duration.

INSTRUMENTATION. *Instrument decay* occurs when observers become bored, fatigued, or more skillful in measuring the dependent variable over the course of the experiment (Campbell, 1957). In an example to be

described later, observers at times manifested all three of these problems. From behind a one-way mirror, each observer recorded the behavior of five hyperactive third-grade students for six hours a day over a three-month period (Chadwick and Day, 1971). A few observers became bored with the task and became lax in their observations. They were observed talking to other observers, reading novels, and filling out crossword puzzles when they should have been watching their assigned students. Other observers became involved with their assigned students and spent time with them outside the experiment. They started to apply different standards in their observations to make these students look good. In the case of both boredom and involvement, the observers undercounted the dependent variable, which was disruptive behaviors. Other observers became more skillful in detecting the behavior they were observing and recorded a higher frequency of disruptions as they became more sensitized, even though independent observations showed that the subject's behavior had not changed appreciably.

STATISTICAL REGRESSION. Statistical regression is the tendency for extreme behavior to be replaced as a general rule by less dramatic behavior. In other words, persons who are statistically deviant on a single test are likely to show up less deviant on a second testing, even if no significant intervention has occurred. The regression effect is a serious concern when subjects are recruited for an experiment on the basis of their exhibiting extreme behavior. For example, subjects selected because they are hyperactive, or have a high rate of delinquent behavior, or have low academic achievement will have a tendency in subsequent tests to drift toward the level of activity that characterizes their normal peers. A researcher who conducts an experiment attempting to reduce the aggressive behavior of hyperactive students or delinquent behavior of teenagers may attribute decreases in such behavior to the independent variable when in reality the effect observed was merely statistical regression.

Campbell and Stanley (1966) suggest that this effect is caused by the larger measurement error at the extremes of a scale. For example, they argue that an extremely high scorer may have had unusually "good luck" (large positive error) and that the extremely low scorer may have had unusually bad luck (large negative error). On the next test or observation the portions of their scores attributable to luck are likely to decrease and the "high" scorer will score a little less high and the "low" scorer a little less low.

DIFFERENTIAL SELECTION. Most experiments compare the behavior of subjects in an experimental group who were exposed to the independent variable with the behavior of those in the control group who were not. It is assumed that the subjects in experimental and control groups were identical before participation in the experiment. However, sometimes sub-

jects are selected for an experimental group in ways that make them different at the outset from persons in the control group. For example, occasionally volunteers are used in an experimental group and nonvolunteers in the control. The differences between volunteers and nonvolunteers noted earlier may be mistakenly attributed to the independent variable.

EXPERIMENTAL MORTALITY. **Experimental mortality** refers to subjects dropping out of the experiment before it is completed. If either the experimental or the control group has a higher rate of mortality than the other group, the differences noted after manipulation of the independent variable may reflect the dropping out of certain types of subjects rather than exposure to the independent variable. Campbell and Stanley (1966) illustrate the mortality effect in a natural experiment testing the relationship between college education and women's physical characteristics. A comparison of the beauty of women entering college with that of those who graduate will generally, they say, reveal that college attendance is "debeautifying." What has happened, however, is that the more attractive women have married and dropped out of college at a higher rate than have the less attractive women.

EXPERIMENTER BIAS. Rosenthal (1966) identified several characteristics of the experimenter that may bias the findings and threaten internal validity. He reviewed research demonstrating that the sex, age, race, and religion of the experimenters may influence subjects' behavior. Personality characteristics displayed by the experimenter such as anxiety, approval, hostility, authoritarianism, dominance, and intelligence have also been observed to alter subjects' behavior in an experiment.

Design We will now review six different experimental designs that attempt to maintain internal validity by controlling for different combinations of the extraneous factors discussed above. There are many other experimental designs, but these six illustrate the major types.

ONE-SHOT CASE STUDY. The one-shot case study involves studying a single group once, following its exposure to the independent variable as noted in Illustration 7.8. In the example of the class length-academic achievement experiment, we would organize the entire ninth grade to one level of the independent variable, say 80 minutes, and then observe the subjects' academic achievement at the end of the school year. It is a misnomer to call this design an experiment, as it does not allow for any meaningful comparisons. There is no pretest observation of the dependent variable to compare with the posttest observation. Nor is there a control group for comparison. Thus, none of the extraneous variables are controlled, and the experimenter cannot make any reliable statements about the relationship between class length and achievement. Campbell and Stan-

ILLUSTRATION 7.7 This cartoon illustrates the potential effect of a personality characteristic of an experimenter on the results of the experiment.



"I think we should ask Zimmer to do those experiments. He's a Capricorn."

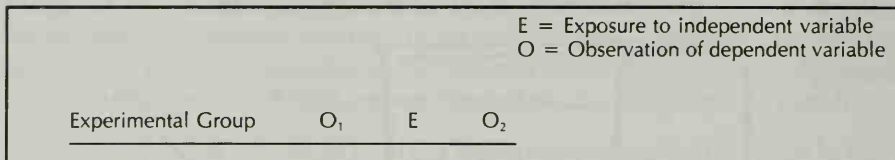
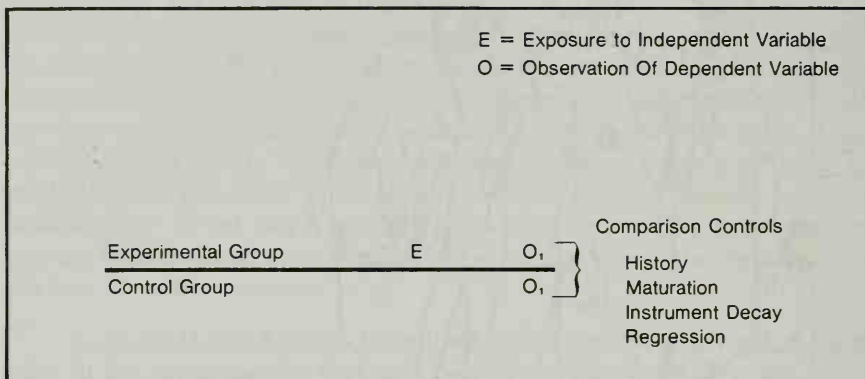
© Sidney Harris. From *All Ends Up*, Los Altos, Calif.: Kaufmann, Inc. Used with permission.

ILLUSTRATION 7.8 One-shot case study design.

			E = Exposure to independent variable
			O = Observation of dependent variable
Experimental Group	E	O ₁	

ley (1966) rejected this design for educational research, and we extend the rejection to all social science experimentation.

ONE-GROUP PRETEST-POSTTEST DESIGN. This design, shown in Illustration 7.9, is occasionally used in the social sciences, but it is susceptible to error from all of the extraneous variables mentioned above. It is

ILLUSTRATION 7.9 One-group pretest-posttest design.**ILLUSTRATION 7.10** Static-group comparison design.

not a much better design than the one-shot case study even though it has a pretest observation that can be compared with the posttest observation. In the class length-achievement example, the students would take the achievement test (CAT) and then for the entire school year would attend classes at one level of duration, say 80 minutes. At the end of the year the students would again be given the CAT. Unfortunately, history would not be controlled, because unknown events rather than class length might have caused any observed differences in achievement. Maturation could also be responsible, for the experiment would last nine months. A test effect is another reasonable explanation for any differences observed even if different versions of the test were given. Since this illustrative experiment would utilize a standardized achievement test with normal students as subjects, instrument decay or statistical regression would probably not be a serious problem. Even so, the design does not allow the researcher to determine whether instrument decay or statistical regression were responsible for any of the observed change.

STATIC-GROUP COMPARISON DESIGN. The static-group comparison design, shown in Illustration 7.10, involves the use of a control group. The usual strategy is to randomly assign subjects to either the experimental or control group, thus controlling for the differential selection effect. In the example of the class length-academic achievement experiment, students would be randomly assigned to one of six experimental groups or to the

control group. The experimental groups would attend classes of specified duration, say, 20, 40, 60, 80, 100, and 120 minutes for the school year. The control group would attend classes of whatever duration is presently in use, probably 50 minutes. In May the students would take the California Achievement Test, and the scores for each group would be compared. The class period whose students had the highest academic achievement scores would be identified as reflecting the class-duration producing the maximum level of academic achievement.

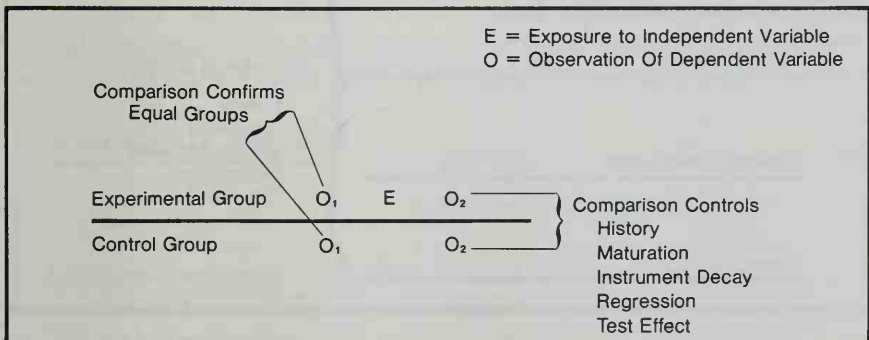
The static-group design controls for history, as both the experimental and the control group are generally influenced by such factors. Maturation is not a problem as both groups mature at the same rate. There is no pretest, so a testing effect is not possible. Instrument decay should not be a problem, as the observations of the experimental and control groups are made at the same time with the same instrument. Any statistical regression effect would be the same for both groups also. All things considered, the static group is a good design controlling most of the extraneous variables.

The major weakness of this design is the assumption that the experimental and the control groups were identical at the beginning of the experiment. Since no pretest observations were made, this remains an assumption. If a reasonably large number of subjects are used and assignment to either experimental or control group is truly random, then the assumption of equivalence at the outset is not unreasonable.

Another potential problem is that this design is susceptible to differential rates of mortality or dropping out. However, if the experimenter determines that both groups had a low rate of mortality, then mortality would not bias the findings. In sum, the static-group design is an adequate experimental design if large samples of randomly assigned subjects are involved and if little mortality occurs.

PRETEST-POSTTEST CONTROL GROUP DESIGN. The pretest-posttest control group design is shown in Illustration 7.11. As was true of the static-group design, it is important that subjects be randomly assigned to

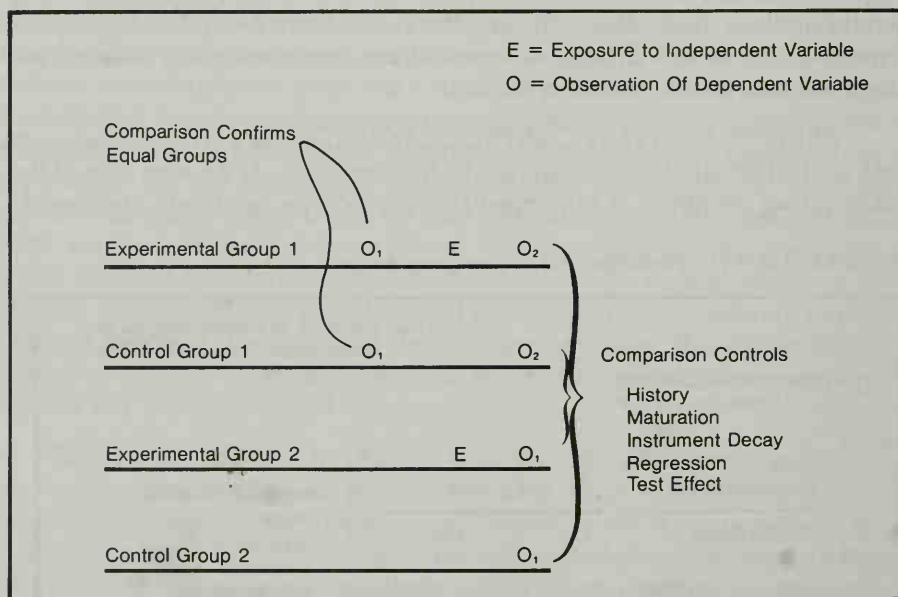
ILLUSTRATION 7.11 Pretest-posttest control group design.



the experimental and control groups. Whether random assignment created groups with identical levels of the dependent variable and other relevant factors can be verified. In the class length–achievement experiment, ninth-grade students would be randomly assigned to one of six experimental groups or one control group as discussed above. They would then be administered the CAT to learn if the groups all had similar, if not identical, levels of achievement. The students would attend school for nine months in one of the class-duration groups and at the end of the school year the CAT would again be administered. The class duration of the group with the highest academic achievement would be assumed to be the duration fostering the highest academic achievement. As is shown in Illustration 7.10, history, maturation, testing, instrument decay, statistical regression, differential selection, and mortality are all controlled for. Differences in the posttest observations between the experimental and control groups can reasonably be assumed to reflect the influence of the independent variable.

SOLOMON FOUR-GROUP DESIGN. The Solomon Four-Group design is the Cadillac of experimental designs, enjoying high prestige, and justifiably so. The basic design, as shown in Illustration 7.12, involves randomly assigning the subjects to one of two experimental and two control groups. The class length–achievement experiment would require randomly assigning subjects to twelve different experimental groups (two for each class length) and two control groups. As noted in the illustration, all of

ILLUSTRATION 7.12 Solomon four-group design.



the extraneous variables are controlled for. The testing effect is given special attention and the magnitude of such an effect can be determined by comparing the posttest observations of the two control groups and making the same comparison between the two experimental groups. An additional strength of this design is that it provides two tests of the hypothesized relationship, thus enhancing the generalizability of the findings. If the posttests for both experimental groups reveal the same effect, then greater confidence is given to the inference that the independent variable is what caused the change in the dependent variable.

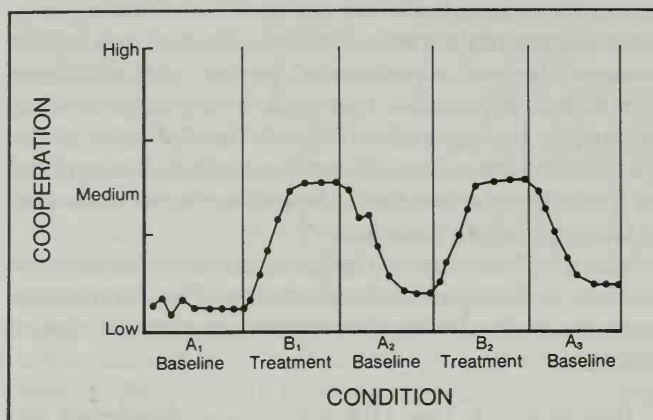
The reason the Solomon Four-Group design is not used more often is that it is more expensive and requires many subjects. The extra experimental and control groups usually double the required number of subjects and the cost of the study.

ABA DESIGN (TIME SERIES). The ABA design was developed primarily by behavioral psychologists to test principles of social learning theory (behavior modification). As is shown in Illustration 7.13, the usual strategy is to observe the baseline rate of a particular behavior during the A-period. Observations are generally continuous. For example, in an experiment testing the effect of reinforcement on cooperation among five-year-old children, the number of times each child offered to share a toy during a 30 minute A-period would be observed as the baseline rate. It is important to maintain the A-period until the dependent variable is occurring at a reasonably consistent rate, called a *steady state*, even if it means continuing beyond the 30 minutes. The experimental treatment is then imposed during the 30-minute B-period. In the example, a specific number of M&M candies would be given to a child following each act of cooperation. The B-period may also have to be extended beyond 30 minutes until a steady state of cooperation is achieved. Finally, in the 30 minute A₂-period, reinforcement would not be given and the number of cooperative acts would again be carefully observed. In the hypothetical cooperation experiment it is expected that a fairly low level of cooperation would be observed in A₁ and would increase and level out during B (see Illustration 7.14). During A₂ cooperation would probably drop off to approximately the original baseline rate. If a B₂-period was added, cooperation probably would quickly increase to the level formerly achieved during B₁. Sometimes the sequence of experimental-control periods continues for a number of cycles.

ILLUSTRATION 7.13 ABA design.

A = Baseline				
B = Exposure to independent variable				
Experimental and Control Group	A ₁	B	A ₂	

ILLUSTRATION 7.14 Hypothetical relationship between reinforcement and cooperation.



The ABA design is most useful in testing relationships that are reversible, as in the cooperation example above. It is argued by those using the ABA design that the subjects act as their own control group, so that the experiment is comparing at least two control groups (A₁, A₂) to one experimental group (B). When the design is extended to include more than one B-period, it then includes several control and experimental groups. The extraneous factors of history, maturation, fatigue, instrument decay, regression, and selection are controlled, as they are identical for the subjects in the A- and B-periods.

A final step in the development of an experimental design is to ensure that other extraneous factors, such as time of day, location of the experimental room, dress of the experimenter, the time when the experimental and control groups are tested, and so on, do not bias the results. Common sense suggests that running all the experimental groups at 8 A.M. and all the control groups at 8 P.M. would probably introduce some bias. A much better strategy would be to alternate the schedule so that half of the control and experimental groups were conducted in the morning and half in the evening.

Debriefing

It is essential that a debriefing session be built into the research design. Its primary purpose is to correct any deception practiced upon the subjects. For example, if subjects were given false scores on tests or learning activities, they need to understand that the scores were manipulated for the purposes of the experiment. Otherwise some subjects might suffer loss of self-esteem as a result of their participation. Any feelings of hostility or anxiety generated by the experiment also need to be reduced.

Another important function of the debriefing is to determine whether the subjects interpreted the experimental situation as the experimenter intended. Occasionally, a subject will see through the deception or completely misconstrue what happened and thus not be exposed to the independent variable in the intended way. Also, helpful insights into how and why subjects behaved as they did emerge from the debriefing.

Pilot Study

Generally it is a good idea to conduct a pilot study of the entire experimental procedure. A few subjects should be trained, run through the experiment, and debriefed. Analysis of the pilot study reveals problems of design, equipment failure, ambiguous instruction, and other things at a stage in the research where corrections or adjustments are still possible.

Analysis and Report

No research project is completed until the data have been prepared for analysis, the appropriate statistical tests have been computed, the results have been interpreted, and a report has been written. Unless a written report of the experiment is available to other researchers, someone else will have to waste time and resources to retest the hypothesized relationships.

VI. ILLUSTRATIVE EXPERIMENTS

Milgram's Obedience to Authority Experiment

Statement of research problem A series of laboratory experiments conducted by Milgram (1963, 1965, and 1974) was designed to test hypotheses relating several independent variables to obedience. The specific experiment to be reviewed here tested the relationship between the proximity of a victim and the subject's obedience to commands to injure the victim.

Operational definitions Milgram was very creative in the measure of obedience that he developed. The subject was instructed to shock a confederate at specific intervals, using a simulated shock generator. The shock generator, located in front of the subject, had an instrument panel with thirty switches. The switches were labeled with the level of voltage associated with each and ranged from 15 to 450 volts. In addition, the thirty switches were divided into seven sets of four switches and a set of two switches. The following eight labels on the shock generator stressed the pain of the punishment: "slight shock," "moderate shock," "strong shock,"

"very strong shock," "intense shock," "extremely intense shock," "danger: severe shock," and "XXX." The subjects were given a 45-volt shock to provide a frame of reference as to how painful the shocks were.

Obedience was defined as the number of switches the subject threw, supposedly shocking the confederate. When the subject balked at giving another shock, the experimenter gave one final command. "You have no other choice, you must go on" (Milgram, 1965: 60). If the subjects still refused to administer the shock, the experiment was terminated and obedience was operationally defined as the number of switches thrown to that point.

The independent variable, immediacy of the victim, was operationally defined as one of four conditions in which the pain cues of the victim were brought closer to the subject. In the *Remote Feedback Condition*, the subject and confederate were in different rooms. At 300 volts the confederate pounded on the wall, and after 315 volts, he no longer communicated with the subject. The *Voice Feedback Condition* was identical except that tape-recorded complaints could be heard through a slightly ajar door. Starting with 75 volts, the confederate grunted and groaned. At 150 volts, he demanded to be released from the experiment. At 180 volts, he cried out that he could no longer endure the pain. At 300 volts, he refused to participate any further in the experiment.

The third condition, *Proximity*, placed the subject and the confederate in the same room less than two feet away. Cues were provided so that the confederate complained at the appropriate time. The final condition, *Touch-Proximity*, required that the subject push the confederate's hand back down on the electrode. At 150 volts, the confederate bridged his fingers so that the palm of his hand was raised off the electrode, even though his arm was strapped down to the chair. The subject was ordered to throw the switch and then reach over and crush the confederate's hand back into contact so that he received the shock.

Selection of subjects One hundred sixty subjects were randomly divided into four experimental conditions. Adult males, age 20 to 50, were recruited in the New Haven, Connecticut, area by a newspaper ad. The ad requested volunteers to participate in an experiment concerning learning and memory, and they were promised \$4 plus carfare for one hour of their time. A good age distribution was obtained: 20 percent of the subjects were in their 20s, 40 percent in their 30s, and 40 percent in their 40s. A wide social class distribution was also obtained, as 40 percent were blue-collar workers, 40 percent were white-collar workers, and 20 percent were professionals. Although not a random sample of American males, this group of subjects was much more representative than a sample of male college students would have been.

Experimental design One naive subject and one confederate, who the subject thought was a fellow subject like himself, performed in each

experiment. The experiment was explained to them as testing the relationship between punishment and learning. They were instructed that one of them was to be the teacher and the other the learner. They drew slips of paper out of a hat to determine who would be which. The drawing was rigged so that the subject was always teacher and the confederate always the learner. The confederate-learner was strapped into an electric chair, and the subject was given a 45-volt shock to provide some awareness of the level of pain associated with even a low-voltage shock.

The task involved the teacher's reading a series of word pairs, such as dog-cat or desk-house. Then he read the first word of a pair and four other words. The learner's task was to remember the correct word to complete the pair. He communicated his choice by pressing one of four switches, which lit up a response on an answer box located on top of the shock generator. The subject-teacher observed the lights and knew whether the learner had made the correct response and, if not, shocked him. The learner was trained to miss two out of every three pairs.

The experimental design was a modified static group design (see Illustration 7.15). The subjects were randomly divided into four groups. No initial pretest observation of obedience was made. The subjects were exposed to the independent variable, that is, different degrees of proximity of the victims. The level of obedience was then observed.

Debriefing Milgram realized the importance of reducing the extreme feelings of tension, anxiety, and guilt generated by participation in the

ILLUSTRATION 7.15 Milgram's obedience experimental design.

	E = Exposure to demands for obedience O = Observations of delivering shocks - - - = Exposure and observations occurring simultaneously
Experimental Group 1: Remote Condition	E - - - O
Experimental Group 2: Voice Feedback Condition	E - - - O
Experimental Group 3: Proximity Condition	E - - - O
Experimental Group 4: Proximity-Touch Condition	E - - - O

experiment. A “friendly reconciliation” between the subject and confederate took place after the experiment, and the deception about the level of shock and the confederate’s pain was explained.

Pilot study At least one pilot experiment was involved in the development of the design described above. The pilot experiment had the subject and confederate in adjacent rooms, connected by silvered glass so that the confederate could be dimly perceived by the subject. No verbal or visual pain cues, other than the words on the shock generator, were used. The initial test of 40 subjects (Milgram, 1963) revealed that a very high percentage of the subjects obediently administered all of the shocks. Voice and visual pain cues were added to test their effect on obedience and to provide evidence on whether levels of obedience varied when such cues were present.

Analysis and report The data were carefully analyzed, and several different reports were prepared (Milgram, 1963, 1965, and 1974; and Elms and Milgram, 1966). A sample of the rather startling results are presented in Illustrations 7.16 and 7.17. These findings stunned the general public as well as social scientists. No one imagined that 30 percent of a sample of adult males would obey an experimenter to the point of delivering a 450-volt shock by personally pushing the victim’s hand on an electrode. Comparisons were made to the “obedience” of German soldiers in carrying out atrocities conducted during World War II.

The extreme tension created during the experiment and guilt afterwards would make it very difficult for a similar experiment to be approved today by an Institutional Review Board. The subjects certainly did not give

ILLUSTRATION 7.16 Percent of subjects who obeyed experimenter to deliver all 30 shocks in four proximity conditions.

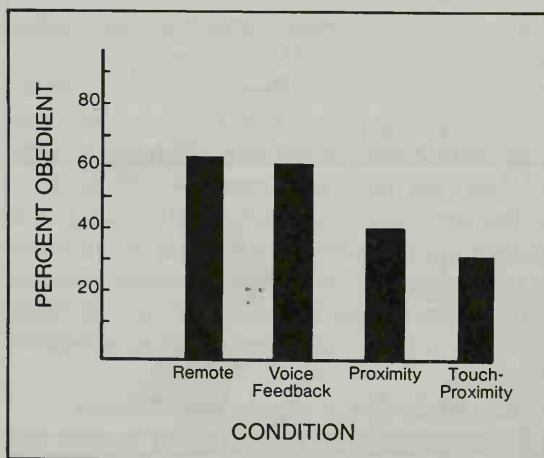
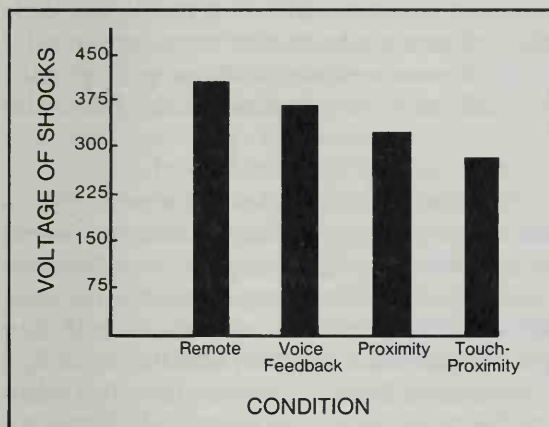


ILLUSTRATION 7.17 Mean maximum shock delivered in four proximity conditions.



informed consent. There was even the potential of a heart attack during the experiment or suicide afterwards. The powerful tension that subjects experienced and its effects on them are illustrated by an observer's comment:

I observed a mature and initially poised businessman enter the laboratory, smiling and confident. Within twenty minutes, he was reduced to a twitching, stuttering wreck, who was rapidly approaching a point of nervous collapse. He constantly pulled on his ear lobe, and twisted his hands. At one point, he pushed his fist onto his forehead and muttered: "Oh God, let's stop it." And yet, he continued to respond to every word of the experiment, and obeyed to the end. (Milgram, 1963: 377)

The question is whether the knowledge coming from the experiment was worth the price paid by the subjects. If one believes the use of deception to create extreme stress in naive subjects is wrong, then the study does not justify the costs. For others who believe society needs to know about the very real danger of excessive obedience to authority and who know that the debriefing diminished the injury to the subjects, the knowledge outweighs the costs. The subjects themselves seemed to feel that the experiment was worth the price. Several years later 84 percent said they were glad to have been in it, 15 percent were neutral, and only 1 percent were sorry. Nearly three-fourths, 74 percent, felt they had benefited from the experience (Samuel, 1975: 58).

Behavior Modification of Underachievement and Disruptive Behavior

Statement of research problem A field experiment conducted by Chadwick and Day (1971) tested the effects of tangible and social reinforcement on disruptive classroom behavior and academic achievement of

children of elementary school age. The experiment was designed to reduce disruptive behavior and increase academic behavior using highly reinforcing tangible reinforcers. Another objective was to pair the tangible with the social reinforcers to determine if, via classical conditioning, the social reinforcers would acquire the reinforcing characteristic of tangible reinforcers.

Operational definitions A modified version of the Becker et al. (1967) observational system was used to count disruptive acts. Observers were trained to record five classes of behavior. The first was *gross motor behavior* which included standing up, leaving a desk, jumping on chairs, moving desks into the aisle, and similar activities. *Loud noise* was the second category of disruptive behavior and included stamping feet, whistling, yelling, banging on desks with a rock or other hard object, and the like. The third category was *verbal aggression*, including name calling, ridicule, and threats. *Behavioral aggression*, the fourth category, included kicking, punching, biting, hitting with thrown objects such as rocks or pencils, and tearing up another student's work. Sometimes the subjects moved so fast that the number of individual blows could not be counted and observers then recorded sequences of aggressive acts. Thus, if a student hit another two or three times in rapid succession before the other could counterattack or retreat, the behavior was recorded as a single aggressive act. The final category was *offensive symbolic behavior*, primarily hand gestures which aroused anger in the recipient and laughter in other students. Whether the verbal, behavioral, or symbolic aggressive act was directed at a peer or the teacher was also recorded.

Academic behavior was measured in several ways including the California Achievement Test (CAT), time worked, rate of work, and accuracy of work. The CAT was administered to the students the first week of the experiment and again during the last week. *Work time* was measured using two electric clocks for each subject. When the teacher gave an assignment, the student's total-time clock was turned on and remained on until she called for the assignment to be handed in. The student's work-time clock was turned on and off as he or she was observed to be working. The most difficult timing decision was encountered when the student stopped working to stare out of the window as if he or she were contemplating the task. A one-minute period was allowed for the student to return to the task before the clock was stopped. Work time was calculated by dividing the minutes worked by the number of minutes allowed to complete the assignment. *Rate of work* was measured by dividing the number of units of work (math, spelling, reading, etc.) by the number of minutes actually worked. The curriculum was designed using roughly comparable units for a given subject so that twenty units of English on one day was approximately equal to twenty units on another day. *Accuracy* was determined by grading a student's work and calculating the percent done correctly.

Each of the observers had responsibility for five children, whom they observed during the entire twelve-week experiment. Reliability was checked by two extra observers, one for disruptive behavior and one for work time. Each day the reliability observers recorded the behavior of five randomly selected children. Their observation were compared with those of the regular observers. Over the twelve weeks reliability averaged 86 percent.

The independent variable, *tangible reinforcement*, was measured with a "point book" which could be used to purchase lunch, field trips, and goods from the school store (where reinforcers such as candy, gum, records, games, baseball equipment, makeup, and similar objects were available). During the experimental period each student was given a point book with three different colored pages in which points for appropriate social and academic behavior were recorded. The first 25 points a student earned each day were placed on a green page to pay for lunch. We were prepared to deny a student lunch if they did not have sufficient points but during the entire experiment not once did a student fail to earn lunch.

Once lunch was earned, a student could choose whether to record points earned on yellow pages to spend in the store after school, or to save points on red pages toward the weekly field trip. Friday afternoon was spent on fishing, swimming, boating, and camping trips and required an accumulation of many points. Quite frequently students failed to earn or save sufficient points, and were taken home at noon on Friday.

A careful schedule of awarding points was worked out. Students earned 20 points for quietly entering the classroom and starting to work within three minutes. They earned 20 points for every 20 minutes of continuous work. Ten points were awarded for continuing to work while a fight or disturbance was raging nearby. Raising of a hand for the teacher's attention and asking permission to leave one's seat was worth 5 points. When disruptive behavior occurred, the teacher quietly but firmly asked the student three times to stop. If the third request was ignored, 10 to 50 points were ripped out of the student's point book.

Academic behavior was awarded points on a schedule that favored accuracy (67 percent) rather than speed (33 percent). When an assignment was completed, it was collected, immediately corrected, and a list of points earned by each student given to the teacher, who immediately gave them to the students so that the delay between work and reinforcement was never more than half an hour.

Finally, the teacher was trained to pair social approval or praise with the tangible reinforcers. An observer recorded the teacher's behavior into three categories: approval, disapproval, and instruction. Only behavior toward a specific student was recorded; comments made to the class as a whole were ignored.

Selection of subjects Elementary teachers in grades two through five in three contiguous school districts were asked to identify minority students

who were having behavioral problems and were "underachievers." Nearly 200 nominations were received from which the 30 most extreme cases were selected.

The sample included 19 Mexican American and 11 black children. Five Mexican American children dropped out of the experiment because their families moved to another community seeking employment. This left 25 students: 16 males and 9 females. They had a mean GPA of 1.47 (D) for the preceding school year and were 1.5 years behind the norm for their age on the CAT.

Experimental design The experiment lasted twelve weeks (June-August) and was conducted in a modern school building. A large classroom was partitioned in half with one-way glass. An observation booth was constructed with tables and stools of different heights so that all observers had an unobstructed view of the classroom. Microphones were hidden in the classroom so that observers could hear all student conversation. The teacher was fitted with an internal hearing device through which the experimenter could communicate to her via a walkie-talkie without the students being aware.

Classes were conducted five days per week from nine in the morning to three in the afternoon. Transportation to and from school was provided. A certified woman teacher conducted classroom activities around a typical curriculum. The observers and equipment were explained to the students as student teachers learning to teach by watching the classroom. This seemed to satisfy all the curiosity about the mirror, equipment, and observers.

The design was an ABA design. A_1 was a three-week baseline when the teacher conducted the class in a conventional way. Students received lunch and field trips and could select articles from the school store up to a specific value each day. The six-week B-period involved the implementation of the point system. Point books were passed out and the contingencies explained. A_2 was basically another baseline, the only difference being that the teacher was encouraged and even instructed via the hearing aid to administer social reinforcement.

Debriefing Debriefing was not as easy as might be expected. Most of the students pleaded to remain in our school rather than return to the public school that was opening the following week. The researchers found it difficult to convince the students to leave the experimental school with a happy attitude.

Finally, parents and the students' public school teachers for the coming year were invited to a conference. Each students' progress was reviewed and suggestions were made on how teachers and parents might maintain the improved academic performance.

Pilot study The first week of the experiment was designated as a pilot study to test the experimental procedures. Several serious problems in what we thought was a very good research design emerged. First, on the opening day of school students literally threw the teacher out of the classroom. She physically could not control them. In addition, the fighting threatened the health of the students, and the equipment, supplies, and the classroom itself were being destroyed. We decided to add a large male graduate student as a teacher's aide. The teacher and aide lasted the rest of the morning but told us at noon that they could not handle the situation. Accordingly, a second teacher's aide, an older male graduate student with several years of elementary school teaching experience, was also sent into the classroom. The two aides were also given hearing aid receivers so that the experimenter or other observers behind the one-way mirror could warn them of impending danger. All three learned to follow warnings without question. If the experimenter told one of them to duck or to move to the right, they did so immediately, usually successfully avoiding thrown objects.

The second major problem was that the students were continuously escaping from the classroom. The classroom had an outside door which could not be locked because of fire regulations. About twice a day three or four students would leave the classroom and race through wheat fields surrounding the school. Members of the research staff had to chase them down and physically drag them back. The worst escape experience happened when a Mexican American boy threw a pencil which punctured the cheek of another student. He became frightened and ran. Five black students appointed themselves vigilantes and raced out to capture him. The student pursuers and a graduate student chased him off the school grounds into a public elementary school, out of that school into a bowling alley, and finally the chase ended under the produce counter of a local supermarket. We finally solved the runaway problem by having staff members assigned to hold the door shut for a few days.

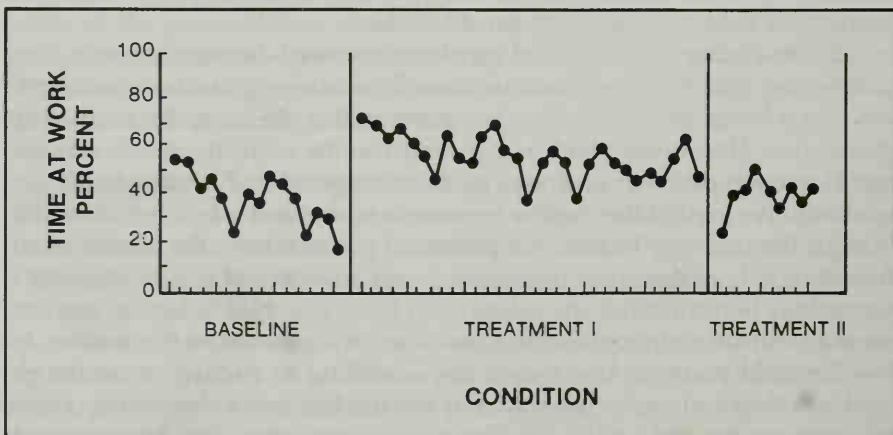
Several other unanticipated problems occurred during the study. The most vexing problem, one that threatened the integrity of the experiment, was that a black girl kept telling her parents that we were discriminating against her. Her mother had had several run-ins with the public schools over the same problem and was quick to respond to her daughter's accusations. We invited the mother to come to school and watch her daughter through the one-way mirror. We promised not to inform the teacher, and she not to tell the daughter. The mother was so incensed at her daughter's pugnacious behavior that she asked us to bring the child to her behind the one-way mirror so that she could beat her. We persuaded the mother to give the child points at home each day according to a schedule we developed and they were to be redeemed by the teacher the next morning. From that time on the girl's behavior improved remarkably. The problem was

that the improvement probably was not the consequence of our reinforcement contingencies but rather was due to the mother's reaction. The following year the girl did so well in public school that her former peers persecuted her and she finally had to transfer to another school in a neighboring town. On the one hand, the mother's involvement had a tremendous positive impact on the life of the student; on the other hand, it confounded the experimental effects and we were forced to eliminate the data on that student from the analysis.

Analysis and report The data were prepared for analysis and a computer file was created. Following appropriate analysis a final report was written and submitted to the funding agency and an article describing the study was published in an appropriate journal (Chadwick and Day, 1971). A sample of the results obtained is presented in Illustration 7.18. Note that work time in the A_1 period (baseline) averaged 39 percent and was declining. This increased to 57 percent in the B period (Treatment). In addition, both rate and accuracy of work increased significantly during the experimental period. However, the use of social reinforcers in the A_2 period (Treatment II) was not very successful in motivating and maintaining work, as the rate for this period was only 2 percent higher than in A_1 .

Disruptive behavior occurred at a rate of 37 disruptive acts per subject per hour during the baseline period. Thus the 25 students averaged 925 disruptive actions every hour. This decreased to only 14 per student each hour during the experimental period or 350 for the class as a whole. More important, the most disruptive and aggressive behaviors such as fighting, yelling, and throwing things almost completely disappeared. Most of the disruptive activity recorded during the experimental period was whispering

ILLUSTRATION 7.18 Percent of time at work during A_1 , B, A_2 periods.



between students. Such behavior probably persisted because it was ignored by the teacher in rewarding or taking away points. During the experimental period one of the two teacher's aides was removed from the classroom and the other remained to assist students in their individual work while the teacher conducted oral reading groups.

Guaranteed Income Experiment

Statement of research problem The alleviation of poverty has received considerable attention from public officials and has resulted in the emergence of the present-day proliferation of public assistance programs. A "War on Poverty" was waged unsuccessfully during the 1960s and since then each presidential administration has called for welfare reform. In recent years a guaranteed income or negative income tax has been seen by many as the solution to poverty in the United States. To test the effects of a guaranteed income the Office of Economic Opportunity (OEO) initiated the Guaranteed Income Tax Experiment and the Department of Health, Education and Welfare (HEW) continued it when OEO was phased out.

The income maintenance experiment, costing over \$100 million, is the largest, most expensive social experiment ever conducted in the United States (Neubeck and Roach, 1981). This field experiment tested the effects of a guaranteed income on poverty-related behavior including work effort, family stability, consumer behavior, educational effort, health and health care, and selected personality factors. Subjects were not invited into the laboratory for a one-hour experiment. Rather, they were asked to participate in a guaranteed income program for a lengthy period of time, some for three years, others for five, and a few for twenty years (Neubeck and Roach, 1981). Participation in the experiment had and continues to have a very real impact on the daily lives of the subjects. The experiment started in New Jersey in 1967 and was later expanded to rural North Carolina; rural Iowa; Gary, Indiana; Seattle, Washington; and Denver, Colorado. We will focus on the Seattle/Denver experiment, as it is the largest and has existed longer than the others.

Operational definition Space limitations prevent a description of all the operational definitions of the poverty-related behaviors studied. We will examine work effort and family stability because they have received the most public attention (Watts and Rees, 1977; Hannan, Tuma, and Groeneveld, 1977; and Moffitt, 1981). Work effort and family stability were measured by self-report in interviews conducted three times a year. The participating families were asked to tell about their efforts to find work, and how much they had worked in the previous four months. Family members were also asked about their marital status to determine rates of family breakup. Families who separated or divorced frequently reported

to experimenters between regular interviews so that benefits received in the experiment could be continued. In order to examine the persistence of changes in behavior, the interviews were continued for two years after the subjects' participation in the experiment had ended.

The concept of guaranteed income is simple. A family is guaranteed an income of a specific level and if they fail to earn that amount, they receive a cash grant bringing the family income up to the guaranteed level. However, when the experiment was set up, difficult decisions had to be made about what level of income to guarantee and how much of the income subsidy a family would lose if its members also earned income.

The decision was made to have three levels of guaranteed income: 90 percent, 125 percent, and 140 percent of the current poverty level. Each of the groups receiving these levels of guaranteed income were then subdivided into four subgroups with different tax rates to reduce benefits when members earned income. Two consistent tax rates were implemented so that one group paid 50 percent and the second 70 percent of money they earned in taxes (i.e., their government subsidy was reduced this amount). The two tax rates gave the families different incentives to work and exceed the guaranteed level, as they could keep 50 or 30 percent of the extra income depending on which group they were in. The other two subgroups operated under a sliding tax scale which increased the tax rate as family income increased. One sliding scale started at 70 percent and the other at 80 percent, and for every thousand dollars the family earned, the tax rate increased by 2.5 percent. Thus, the independent variable of the guaranteed income was manipulated by varying situations of the different subgroups, ranging from the control group through the twelve experimental groups each with different levels of guaranteed income and different tax rates on earned income.

Selection of subjects Low-income areas in Seattle and Denver were canvassed for family heads between 18 and 58 years old. To qualify for participation, families had to have an income under \$9,000 with a single wage earner, and under \$11,000 if there were two wage earners. A very large sample of 4,800 families was selected. It included both one- and two-parent families, and many of the one-parent families were headed by females. The sample also had a significant number of Hispanic and black families. The large sample increased the probability that the families selected were reasonably representative of families living in poverty in urban America. The experiments in rural North Carolina and rural Iowa were conducted to test the effects of a guaranteed income among nonurban families.

Experimental design The experimental design was a modified pretest-posttest control group design. The 4,800 families were randomly divided

into one control group and twelve experimental groups (see Illustration 7.19). Nearly half of the families, 44 percent, were placed in the control group. Three out of four experimental subjects participated in the three-year experiment and the remainder in the five-year experiment. A few families in Denver are involved for a twenty-year period (1971–1991). Subjects knew the length of the experiment when they agreed to participate.

Pretest data about work effort and family stability were collected during the process of screening the families for participation. Observations were obtained three times a year during the experiment and for two years after it was terminated.

Debriefing We have been unable to find comments in the published reports, but we suspect that many subjects resisted being phased out of the experiment. Presumably, many subjects in the three-year experiment were aware of the five-year and twenty-year experiments and argued that they should be allowed to remain in the program. In some cases, readjustment to life without the guaranteed income must have been a traumatic experience.

Pilot study The original experiment in New Jersey served as a pilot study for the projects in the other sites. The pilot study used a somewhat restricted sample: only families with a male head were selected. The results were criticized as not being generalizable to female-headed families, which are a sizable segment of the poverty population (Neubeck and Roach, 1981). Families headed by women were included in the Seattle/Denver samples. Another criticism was that the three-year period of the New Jersey experiment was too brief and did not allow the guaranteed income to have significant impact. Therefore, some of the Seattle//Denver subjects participated for five years and a few will participate in the program for twenty years. As a result of the lessons learned in the New Jersey experiment, the Seattle/Denver experiment is a much better designed project than it would otherwise have been.

Analysis and report The data have been and continue to be analyzed by many social scientists from several disciplines including economics, sociology, and psychology. Several scholarly reports and books on certain aspects of the experiments have appeared. The interested reader is referred to the references in Neubeck and Roach (1981). In addition, the popular press has widely disseminated the rather startling findings (*New York Times*, 1978, *Newsweek*, 1978, and *Fortune*, 1978).

Perhaps the greatest surprise to emerge from the experiment was that a guaranteed income program increased separation and divorce.

Consider first the impact on the probability of the dissolution. Our findings

ILLUSTRATION 7.19 Design of Seattle/Denver guaranteed income experiment.

	O = Observation of work effort and marital sta- tus
	E = Participation in guaranteed income program
	----- = Participation and observations occurring simul- taneously
CONTROL GROUP	
	O ₁ E O ₂
Exp. Group 1 90% Poverty Level: 50% Tax Rate	O ₁ - - - E - - - O ₂
Exp. Group 2 90% Poverty Level: 70% Tax Rate	O ₁ - - - E - - - O ₂
Exp. Group 3 90% Poverty Level: Sliding Tax Rate	O ₁ - - - E - - - O ₂
Exp. Group 4 90% Poverty Level: Sliding Tax Rate	O ₁ - - - E - - - O ₂
Exp. Group 5 125% Poverty Level: 50% Tax Rate	O ₁ - - - E - - - O ₂
Exp. Group 6 125% Poverty Level: 70% Tax Rate	O ₁ - - - E - - - O ₂
Exp. Group 7 125% Poverty Level: Sliding Tax Rate	O ₁ - - - E - - - O ₂
Exp. Group 8 125% Poverty Level: Sliding Tax Rate	O ₁ - - - E - - - O ₂
Exp. Group 9 140% Poverty Level: 50% Tax Rate	O ₁ - - - E - - - O ₂
Exp. Group 10 140% Poverty Level: 70% Tax Rate	O ₁ - - - E - - - O ₂
Exp. Group 11 140% Poverty Level: Sliding Tax Rate	O ₁ - - - E - - - O ₂
Exp. Group 12 140% Poverty Level: Sliding Tax Rate	O ₁ - - - E - - - O ₂

implied that, if the entire sample were enrolled in an income-maintenance program with a low support level, the annual probability of marital dissolution would increase 63% for Blacks, 194% for Whites, and 83% for Chicanos over what it would be in the control situation. (Hannan, Tuma, and Groeneveld, 1977: 1206)

This finding had significant political implications, as one of the major selling points of a guaranteed income program was that families would be strengthened because it would not be necessary for the husband/father to leave the home for the family to qualify for welfare benefits. It has been suggested that the guaranteed income program has the unanticipated effect of making husbands and wives financially independent of each other and this independence explains much of the increased family dissolution (Hannan, Tuma, and Groeneveld, 1978).

Negative findings also surfaced in regard to participation in paid employment. The results indicated unequivocally that hours of work were reduced by the guaranteed income program.

The disincentive effects for husbands range from one percent to eight percent. For wives, they vary much more—from almost zero to 55 percent (although the latter figure may be a statistical anomaly). Disincentives of 12 to 28 percent were reported for female family heads in the only two experiments for which estimates are available, Gary and Seattle/Denver. (Moffitt, 1981: 24-25)

In other words, the families receiving guaranteed income reduced the amount of time they were employed as compared with the control group, despite the tax rate incentives built into the program. It had been hoped that the guaranteed income combined with the tax rates would motivate family members to find and keep employment while providing a net or threshold of safety against real need when work was unavailable.

Supporters of the guaranteed income program suddenly found themselves faced with the facts that such a program reduced people's tendency to work, increased their dependence on government handouts, and weakened family ties. Senator Patrick Moynihan, a former supporter of a guaranteed income program, summarized the feelings created by the unanticipated findings when testifying before the United States Senate:

It does not seem likely that the answers will be comforting to those of us who had hoped to replace the existing programs with some form of national income maintenance or negative income tax program. The evidence presented to us last Spring suggested that, far from strengthening family ties, such a reform might further weaken them. Moreover, instead of encouraging work and self-sufficiency, the kinds of plans tested appeared to produce substantial reduction in work effort and corresponding increases in dependence on public subsidy. Ten years ago, we expected quite different outcomes from these tests. We must now be prepared to entertain the possibility that we were wrong. (U.S. Senate, 1978: 2-3)

The guaranteed income experiment is an excellent example of a field experiment that has had significant impact on public policy in American society. The cost in dollars and human suffering of the scientific experiment are small compared with what the costs would have been if the federal government had implemented a guaranteed income program on a national level.

VII. SUMMARY

An experiment involves the researcher's manipulating one or more independent variables and observing the effects on a dependent variable. A laboratory experiment, conducted in a highly structured environment, gives the experimenter excellent control over the relevant variables. A field experiment is conducted in a real-world social environment such as a school, factory, or community and is more realistic than the laboratory experiment, but the experimenter has less control over extraneous factors that may affect changes in the dependent variable. In a natural experiment the researcher allows other forces such as natural events, economic changes, political events, and so on to manipulate the independent variable. The natural experiment benefits from a realistic setting but does so at the cost of reducing the control the experimenter can impose on the experiment.

The advantage of experimental research is the opportunity to establish cause-and-effect relationships which then can be used by political leaders, administrators, and other interested parties to produce desired changes in social institutions and individual behavior. The difficulties or weaknesses of experimental research include problems in recruiting appropriate subjects thus limiting the ability to generalize to larger populations, bringing the real world into the laboratory thus reducing internal and external validity, and some risk of injury to subjects. There are now federal regulations designed to protect human subjects and there is talk of expanding the protections guaranteed.

The stages of experimental research are, first, the development of a clear statement of the research problem. Second is the creation of operational definitions including the experimental tasks subjects will be asked to perform. The third stage is the selection of experimental subjects. Fourth is the development of an experimental design controlling relevant extraneous variables so that the findings can be accepted with confidence. Threats to internal validity (history, maturation, testing, instrumentation, statistical regression, differential selection, mortality, and experimenter bias) can be controlled by an appropriate design. The static-group comparison, the pretest-posttest control group, the Solomon four-group, and the ABA time series designs all control threats to internal validity to some extent. The fifth step in experimental research is to develop debriefing procedures so

that potential injury to subjects is reduced and any deception explained and justified. The sixth stage is to conduct a pilot study to test the experimental procedures and equipment. Finally, the data are collected, analyzed and interpreted, and a report is written.

Milgram's Obedience to Authority experiment is an excellent example of a creative, realistic laboratory experiment. Chadwick and Day's experiment with hyperactive children illustrates a field experiment in a school setting with reasonably good control. The Guaranteed Income experiment is an example of a large-scale, long-term field experiment in two large metropolitan areas.

REFERENCES

- BECKER, W. C., C. H. MADSEN, CAROLE R. ARNOLD, AND D. R. THOMAS, 1967. "The Contingent Use of Teacher Attention and Praise in Reducing Classroom Problems." *Journal of Special Education* 1, No. 3: 287-307.
- CAMPBELL, DONALD T., 1957. "Factors Relevant to the Validity of Experiments in Social Settings." *Psychological Bulletin* 54: 297-312.
- CAMPBELL, DONALD T., and JULIAN C. STANLEY, 1966. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.
- CHADWICK, BRUCE A., and ROBERT C. DAY, 1971. "Systematic Reinforcement: Academic Performance of Underachieving Students." *Journal of Applied Behavior Analysis* 4 (November): 311-319.
- COLEMAN, JAMES S., 1975. "Racial Segregation in the Schools: New Research with Policy Implications." *Phi Delta Kappan* 57: 75-78.
- COLEMAN, JAMES S., et al., 1966. *Equality of Educational Opportunity*. Washington, D.C.: U.S. Government Printing Office.
- DAY, ROBERT C., 1971. "Some Effects of Combining Close, Punitive, and Supportive Styles of Supervision." *Sociometry* 34: 303-327.
- ELMS, A. C., and S. MILGRAM, 1966. "Personality Characteristics Associated With Obedience Toward Authority Command." *Journal of Experimental Research in Personality* 2: 282-289.
- Fortune, 1978. "Some Negative Evidence About Negative Income Tax." *Fortune* 98 (December 4): 145-147.
- HANNAN, MICHAEL T., NANCY BRANDON TUMA, and LYLE P. GROENEVELD, 1977. "Income and Marital Events: Evidence from an Income-Maintenance Experiment." *American Journal of Sociology* 82 (May): 1186-1211.
- 1978. "Income and Independence Effects on Marital Dissolution: Results from the Seattle and Denver Income-Maintenance Experiments." *American Journal of Sociology* 84, No. 3: 611-633.
- HOLMES, COOPER, 1979. "Sample Size in Psychological Research." *Perceptual and Motor Skills* 49: 283-288.
- HOLMES, COOPER B., JUDITH R. HOLMES, and JOHN J. FANNING, 1981. "Sample Size in Non-APA Journals." *The Journal of Psychology* 108: 263-266.

- MCNEMAR, QUIN, 1946. "Opinion-Attitude Methodology." *Psychological Bulletin* 43: 286-294.
- MILGRAM, STANLEY, 1963. "Behavioral Study of Obedience." *Journal of Abnormal and Social Psychology* 67: 371-378.
- 1965. "Some Conditions of Obedience and Disobedience to Authority." *Human Relations* 18: 57-76.
- 1974. *Obedience to Authority*. New York: Harper and Row.
- MILGRAM, STANLEY, and R. SHOTLAND, 1973. *Television and Antisocial Behavior: Field Experiments*. New York: Academic Press.
- MOFFITT, ROBERT A., 1981. "The Negative Income Tax: Would It Discourage Work?" *Monthly Labor Review* (April): 23-27.
- NEUBECK, KENNETH J., and JACK L. ROACH, 1981. "Income Maintenance Experiments, Politics, and the Perpetuation of Poverty." *Social Problems* 28, No. 3 (February): 308-320.
- Newsweek*, 1978. "Welfare: A Surprising Test." *Newsweek* (November 27): 33-39.
- New York Times*, 1978. "Moynihan Says New Studies Raise Doubts on Negative Income Tax." *New York Times* (November 16): A23.
- ROSENTHAL, ROBERT, 1966. *Experimenter Effects in Behavior Research*. New York: Appleton-Century-Crofts.
- ROSENTHAL, ROBERT, and RALPH L. ROSNOW, 1975. *The Volunteer Subject*. New York: John Wiley.
- RUBENSTEIN, CARIN, 1982. "Psychology's Fruit Flies." *Psychology Today* 16, No. 7: 83-84.
- RUBIN, ZICK, 1983. "Taking Deception for Granted." *Psychology Today* 17, No. 3: 74-75.
- RUSH, MICHAEL C., HAMES S. PHILLIPS, and PAUL E. PANEK, 1978. "Subject Recruitment Bias: The Paid Volunteer Subject." *Perceptual and Motor Skills* 47: 443-449.
- SAMUEL, W., 1975. *Contemporary Social Psychology: An Introduction*. Englewood Cliffs, N.J.: Prentice-Hall.
- U.S. Senate, 1978. Hearings on Welfare Research and Experimentation Before Subcommittee on Public Assistance of the Committee on Finance. Ninety-Fifth Congress, Second Session (November 15). Washington, D.C.: U.S. Government Printing Office.
- WATTS, HAROLD W., and ALBERT REES (eds.), 1977. *The New Jersey Income Maintenance Experiment*, Vol. II and III. New York: Academic Press.

Qualitative Research

eight

I. Introduction

II. Strengths of Qualitative Research

Viewing behavior in its natural setting

Depth of understanding

Flexibility

III. Problems of Qualitative Research

The violation of rights of human subjects

Legal, moral, and injury risks

Floundering

Loss of detachment

Reliability

IV. Design of Qualitative Research

Degree of participation

Degree of secrecy

V. Doing Qualitative Research

Selection of a topic

Determining the degree of participation and secrecy

Developing cover and exit stories

*Participation
behavior*

meanings

problems

interactions

discrepancies

Exit

Analysis of data and writing of report

VI. Illustrative Examples

Festinger, Riecken and Schachter's "When Prophecy Fails"

selection of a topic

determining the degree of participation and secrecy

cover story and entry

participation

exit

analysis and report

Humphreys' "Tearoom Trade"

selection of a topic

determining the degree of participation and secrecy

cover story and entry

participation

exit

analysis and report

VII. Summary

I. INTRODUCTION

The term **qualitative research** refers to several different modes of data collection, including field research, participant observation, in-depth interviews, ethnomethodology, and ethnographic research. There are substantial differences among these research strategies, but they all emphasize "getting close to the data" and are based on the concept that "experience" is the best way to understand social behavior. Here is a typical description of qualitative research.

Qualitative methodology refers to those research strategies, such as participant observation, in-depth interviewing, total participation in the activity being investigated, field work, etc., which allow the researcher to obtain firsthand knowledge about the empirical social world in question. Qualitative methodology allows the researcher to "get close to the data" thereby developing the analytical, conceptual, and categorical components of explanation from the data itself—rather than from the preconceived, rigidly structured, and highly quantified techniques that pigeonhole the empirical social world into the operational definitions that the researcher has constructed. (Filstead, 1970: 6)

Getting close to the data implies interaction with the people being studied; learning their culture, including their values, beliefs, behavior patterns,

and language; and attempting to feel or experience their motives and emotions. A qualitative researcher understands social behavior because he or she:

. . . discovers the actor's "definition of the situation"—that is, his perception and interpretation of reality and how these relate to his behavior. . . . Finally, in order for the researcher to come to such an understanding, he must be able (albeit imperfectly) to put himself in the other person's shoes. (Schwartz and Jacobs, 1979: 7–8)

Some qualitative researchers reject the scientific quantitative method, arguing that it is a mistake to imitate the research strategies of the physical sciences, as human behavior is different from the subject matter studied by chemists, biologists, and physicists. The uncritical adoption of the methods of the physical sciences is seen by some qualitative researchers as having produced social scientists "who measure everything and understand nothing" (Filstead, 1970: vii). The development and testing of abstract theories using quantitative models is seen as biasing research and distorting social reality by forcing it into theoretical pigeonholes. Operational definitions and a concern with quantification are seen as impediments which prevent scientists from studying unobservable or "inner experiences." In its extreme form the emphasis on measurement of observable events is seen as removing the social scientist from what he or she is actually trying to understand. Below is a typical statement of this position:

I am questioning the value of highly complex measuring devices that become ends in themselves rather than intermediary tools used to increase the amount of sociological understanding. It is inexcusable to force the research problem into an *a priori* scheme of technical paraphernalia rather than observing it in the context of the empirical world being investigated. (Filstead, 1970: vii)

Not all qualitative researchers take this extreme position; many organize their records around formal theories, and much qualitative research includes careful counting and recording of events. Nevertheless, the objective is to describe social realities from the perspective of the subjects, not the observers. The definition of the situation of the people being studied includes their observable behavior and also their subjective motives, feelings, and emotions (Schwartz and Jacobs, 1979: 5). The concern with subjective definitions reduces the usefulness of theories, models, and measurement, for often the "inner events" are defined as understandable only through a researcher's personally experiencing how the subjects see and interpret their lives. Thus, using the methods of many qualitative researchers, one would study prostitution from the perspective of the prostitute and her client rather than from the perspective of legislators, police officers, judges, social workers, or other outsiders.

Some qualitative researchers say that an accurate understanding of others' perspectives is best achieved if the researcher brings to the research setting as few preconceived theories or ideas about measurement as possible. Rather, the ideal approach is for the researcher to immerse himself or herself in interaction with the research subjects and their surroundings and let the language of description and an awareness of social patterns emerge from deep involvement with the subjects. Gaining such an awareness is not easy. The researcher must suspend personal values, perceptions, and feelings and try to experience the world from the viewpoint of the selected others.

Participant observation is only a gate to the intricacies of more adequate social knowledge. What happens when one enters that gate depends upon his abilities and interrelationships as an observer. He must be able to see, to listen, and to feel sensitively the social interactions of which he becomes a part. He must be able to grow with his experiences. He must question time and time again whether he has perceived enough and whether his understandings are as accurate as he can make them. He must be able to understand his own impact upon the social situation he studies and what influences other participants and the situation have upon him. He must learn to expect a personal sense of culture shock, a confusing and possibly a painful experience, as a symptom that he is bringing new perceptions of situations into focus and that he is becoming able to assimilate those perceptions into his modifying understandings. (Jacobs, 1970: 7)

These principles are illustrated in a study of cab drivers (Henslin, 1972; see also Henslin, 1967, 1968, and 1974), who tried to discover how sufficient trust develops between the driver and his passenger that the driver is confident he will be paid and not harmed.

Henslin began with a common-sense discussion of the potential hazards a cabbie faces and defines "trust" as a situation where "an actor [customer] offers a definition of himself and an audience [cabbie] is willing to interact with the actor on the basis of that definition" (Henslin, 1972: 22). Henslin's description of research methods amounts to a statement that he drove a taxi to gain his information and that the other drivers thought he was "simply another cabbie." No one knew that he was a graduate student in sociology studying the working life of the cab driver. We are not told how long he drove a taxi, how many passengers he carried, how often a passenger refused to pay, or how many times he was robbed. From the qualitative research perspective whether he drove one day or ten years is irrelevant; the important thing is that he experienced the problems and feelings of a cab driver.

As a result of his participant observation, Henslin identified four sets of factors that he used in deciding if a potential passenger was trustworthy. The first was the type of request for service and the time of day. He was much more likely to accept a rider referred from the dispatcher than some-

one who flagged him down, and he was more trusting during the day than at night. The second set of characteristics related to the location of the pickup and included the social class and racial makeup of the neighborhood where the passenger waited as well as the lighting in the immediate vicinity.

The third set of variables was the passenger's socioeconomic status, race, sex, age, and sobriety. Upper- or middle-class individuals, the very old or very young, whites, females, and those who seemed sober or only slightly tipsy were thought to be more trustworthy than were lower-class people, the middle-aged, blacks, males, and those who were obviously drunk.

The final set of characteristics were passenger behaviors, including whether the passenger was seen emerging from the house or apartment from which the call originated, where and how the passenger sat in the cab, and the apparent rationality of his or her behavior. Henslin's insights on how taxi drivers ensure their safety are illustrated in the following quotations from his field notes.

About midnight I was dispatched to an apartment building where I picked up two men who appeared to be in their seventies or eighties. As we drove along I started to count the money in my pocket. Ordinarily every time I accumulated five dollars, over enough to make change for a ten, I would put the excess away to make certain that it would be safe in case of a robbery. I thought to myself, "I should put this money away," but then I thought, "No, these guys aren't going to rob me." It was at this point that I realized that I felt safe from robbery because of their ages. One does not ordinarily think of a robber as being an old man. These men were not spry, they walked with canes; and they didn't look as though they were physically able to rob me. (Henslin, 1972: 26)

It was about 1:00 A.M. I had taken a practical nurse home after her work shift and ended up in part of the ghetto. Since I was next to the stand, I decided to park there. As I was pulling into the space, I saw a man standing at the bus stop which was next to the stand, with his arm held out horizontally and wagging his finger a bit. He was a large black man wearing a dark blue overcoat. He opened the back door of the cab, and my first thought was, "Well here goes! I'm going to be robbed. I'd better turn on the tape recorder and get this on tape!" After he got in the cab he said, "I want to go to Richmond Heights. You know where Richmond Heights [middle-class neighborhood] is? (Henslin, 1972: 28)

Note that Henslin made his decisions on the basis of his own feelings or insights rather than from instruction by other cabbies. He "felt safe from robbery" because of the old men's ages. In the second quotation the reader can almost feel Henslin's relief when the threatening black male asked for a ride to a solid middle-class neighborhood. Also, the second quote revealed that Henslin had a tape recorder available to record conversations with

passengers and apparently the recording was done without the passengers' knowledge or consent.

Henslin's article is a description of how one St. Louis taxi driver handled the risks of his occupation. There are no numbers or statistical analyses. The reader doesn't know how often different types of passengers appeared or how successful the driver was in avoiding trouble. The study is an articulate description of Henslin's experience but there is no evidence that he was a typical driver. It is risky to generalize the findings of qualitative research to other groups or individuals. However, the committed qualitative researcher might respond that generalization of superficial information is not nearly as useful as knowledge about how trust emerges in the experience of a single driver.

This example is close to the qualitative end of the qualitative-quantitative research continuum, as little if any measurement was involved. However, many—perhaps most—qualitative researchers supplement their research experience with counts of specific behavior, structured interviews, and questionnaires. Frequently the observer's senses are supplemented by mechanical devices such as tape recorders, event recorders, video tape recorders, or movie cameras.

A fascinating example of combining qualitative and quantitative research methods was a "sting" operation conducted by the Washington, D.C., Police Department. Participant observation techniques were used to gain access to criminal activity in the area and quantitative methods were used to obtain evidence that would stand up in court. In October 1975 the Washington, D.C., Police Department created Police, FBI, Fencing, Incognito, Inc. (PFFI, Inc.). The police officers who posed as thieves and fences gave the impression that they were connected to the Mafia by wearing loud clothes purchased at second-hand clothing stores, using Italian-sounding names such as Frank Lasagne and Pepi Pepperoni, and trying to talk with an Italian accent. They rented an old warehouse and started to buy (fence) stolen property. The "study" continued for five months and was funded by \$2,000 advanced by an insurance company and \$65,000 from the police budget. This investment was used to purchase \$2.4 million worth of stolen goods and evidence was collected on 10,000 crimes.

The quantitative techniques were very insightful. "Clients" who sold stolen objects to PFFI were served wine in a fine glass which provided exceptionally clear fingerprints. They were also asked for identification, preferably a driver's license and social security number, and in most cases the client quickly produced the documents. All fencing transactions were videotaped and were coordinated with the fingerprints and identifying documents. The police participants hinted that the "Family" had some positions open and interested individuals were asked to fill out a "job application." The application included documenting previous criminal experience. Many unsolved crimes including three murders, ten armed rob-

beries, four bank robberies, and two hijackings were solved by information revealed in the job applications.

It seems strange that the criminals did not see through the charade of PFFI, but this was not the case. The line of potential "clients" waiting their turn was sometimes more than a block long. The creative collection of quantitative data—fingerprints, social security numbers, driver's licenses, videotapes of thieves in the act of selling stolen goods—and the evidence revealed in the criminal histories obtained in the "job" applications resulted in many court convictions. PFFI was so successful that many police departments across the country have replicated it. It represents the imaginative combination of qualitative and quantitative methods in an "applied research" project.

II. STRENGTHS OF QUALITATIVE RESEARCH¹

Viewing Behavior in Its Natural Setting

A major advantage of qualitative research is that it often involves the observation of behavior in its "natural setting." Allegedly, a researcher's understanding is increased because he or she deals with subjects in *their* world, and not in an artificial one created by the researcher. Deutscher (1973: 149) notes that even during an interview, "the situation is so meticulously constructed and carefully managed" by the interviewer that it does not resemble any routine social situation. Becker and Geer (1957) have argued that the biggest difference between qualitative research and other techniques is that the participant observer has a richer experiential context and is thereby sensitized to incongruous or unexplained activities and their implications. Consequently, the researcher is continually forced to revise and adapt theoretical predispositions to make them congruent with the observed behavior and its intuited meanings.

Depth of Understanding

The second advantage of qualitative research comes from its potential "to grasp the native's point of view, his relation to life, to realize *his* vision of *his* world" (Malinowski, 1922: 25). Qualitative research requires the scientist to be directly involved and forces him or her to acquire some degree of understanding about what life feels like and means to the people being studied. Field work helps the researcher to "ground" observations and impressions in a richly elaborated context of the perceived world view and values of the subjects.

¹ Portions of the following two sections were adapted from Freudenberg and Albrecht, 1982.

Flexibility

A third advantage of qualitative research is its flexibility in that it allows a researcher to be "surprised." A questionnaire survey is not likely to provide answers to questions that were not asked, but a qualitative researcher living with a group or in a community may experience and comprehend events and conditions not even considered before the field work began. William F. Whyte (1955: 375) concluded from his study of street corner society that "the researcher will miss important data unless he is flexible enough to modify his plans as he goes along. The apparent 'tangent' often turns out to be the main line of future research."

III. PROBLEMS OF QUALITATIVE RESEARCH

The Violation of Rights of Human Subjects

A difficulty the qualitative researcher must face is the ethical propriety of the research design. This problem is most severe under conditions of high secrecy and total participation. The researcher may obtain revealing and potentially harmful information from subjects without their knowledge and informed consent. The possible use of such data by third parties may have serious consequences for both subjects and researchers. When Laud Humphreys' (1970) research became public knowledge, the St. Louis Police Department requested that he identify the homosexuals observed so that they could be prosecuted. Covert studies of "deviants" have the potential to bring legal sanction or public ridicule upon the subjects. Such potential consequences are a direct responsibility of the researcher, who must decide whether the knowledge to be obtained justifies the risks incurred.

Another ethical issue raised by qualitative research is that the researcher becomes part of the group, at least temporarily, and thus has the potential to change it. Most researchers try to play a neutral or "follower" role but at times find themselves forced into positions where their actions may affect the future of the group or its members. Alfred (1976), in a covert study of the Church of Satan, had to cope with involvement in leadership positions. His feigned conversion to Satanism was accepted as genuine, and he made rapid advancement in the ritual rank, in magical responsibilities, and to appointment to the governing body of the church. In spite of strong efforts to avoid altering the group, Alfred acknowledged that his presence had made a difference in many situations.

Throughout the research, I tried to minimize reactivity, that is, my own effect on the group I was studying. I was generally nondirective in my comments and conversation, demurred in first requests for suggestions or ideas, answered subsequent requests with suggestions made previously in similar

situations by others, and even selected at random pages from books out of which I was asked to read something of my choice. Even in the group's ruling council I was able to avoid undue influence, since it was an advisory rather than a legislative body. . . . Such efforts and devices, however, did not completely solve the problem of reactivity; I often had to choose what ideas to second, since I was generally perceived as a high-status member and since my behavior was interpreted by others as flowing from genuine satanic conviction and devotion to the church rather than as random acts or simple yesmanship. (Alfred, 1976: 84)

Participant observers may not consciously change the groups they study, but by their very presence some change is introduced. The observers must be aware of their responsibility for changes caused either by design or incidentally, and they must realize that their presence in the group being studied creates a different reality from that which would exist in their absence.

The researchers' impact on the group is often strongly felt when a study is completed and the observers "drop out." If they have established close personal ties with group members, the rupture of those ties may be psychologically costly to group members. If the researchers' departure is accompanied by the revelation that they were not what they had seemed to be, feelings of betrayal or exploitation may accompany the sense of loss felt by subjects.

The possible violation of the rights of the individuals studied is a major concern for the qualitative researcher, and usually there is no simple solution to this problem. Ideally, the degree to which group members come to depend on the researchers should be minimized, and the confidentiality of sensitive material about specific individuals or groups must be protected. References to people and places in reports about the study must be altered sufficiently that the anonymity of subjects is maintained.

Legal, Moral, and Injury Risks

Not only does the qualitative researcher risk injury to research subjects, but he or she also risks personal injury. Participant observation in groups engaged in criminal activity may leave the researcher liable to prosecution. Participation in groups whose sanitation practices differ from those to which the researcher is accustomed may encounter disease and long-term health problems. Also, drug or alcohol use accompanying participant observation may have lasting negative consequences for researchers. Similarly, participation in highly emotional groups, such as a satan worshipping cult or a charismatic Pentecostal sect, may affect a researcher's mental health.

Sometimes participant observation may require the qualitative researcher to violate personal moral standards. Some researchers may feel that they can suspend their personal values in the service of science, but

in fact participation in "deviant" behavior may have long-term consequences on self-image and feeling of guilt. The use of alcohol or drugs, participation in premarital or extramarital sexual behavior, the harassment of innocent others, discriminatory acts against members of ethnic groups are among the behaviors that may violate a researcher's personal standards.

The qualitative researcher must choose projects and design their studies carefully so as to minimize the possible violation of deeply held personal standards. Many worthwhile projects are simply not worth the risk of potential damage to self or to others (spouse, children, parents, colleagues) who have some interest in or claim upon the physical and mental health of the researcher.

Floundering

Just as qualitative research allows the researcher to be flexible and to be surprised by unforeseen developments, it also allows the possibility that the data collection will be so unstructured that nothing meaningful emerges from the field work. The probability that extensive field work will prove useless is especially high when the researchers expect to make great progress in a very short time. Qualitative research is inherently time-consuming. For example, after three full years of field work, William F. Whyte (1955) applied for a three-year extension of his Street Corner Society project because he felt he was just beginning to understand what was happening in the group he was studying. Qualitative researchers must keep their data collection focused on central issues and limited in scope. Even then it is a time-consuming task, as a thorough qualitative study of a single family could consume an entire lifetime.

Loss of Detachment

Another problem of qualitative research is "going native," or over-identifying with the subjects of the research. In the very process of being both participant and observer, the researcher may lose objectivity and be socialized into the values and activity patterns of the group. Apart from the relatively rare circumstances where a researcher deliberately exaggerates the positive characteristics of the subjects, the researcher must guard against overidentification. Overidentification has occasionally resulted in the researcher's joining the social group being studied.

Another form of loss of detachment involves confusing part of the social reality for the whole. This problem is most critical in cases where researchers study cultures or groups very different from their own. The danger is that the researchers, more naive than they suspect, believe that they have found "the real group" or a representative segment of the population when in fact they have not. The researchers assume that they understand the entire society when in fact they have focused on a particular

subgroup, often a marginal one, where they feel comfortable. Much qualitative research suffers from an "elitist" bias, for it is the upper-status members of a group who tend to be the official or unofficial spokespersons for the group or community and have great contact with the researchers. Also upper-status informants are more articulate and better informed than other members of the group and researchers are attracted to these better informed subjects.

Reliability

To a greater degree than in other research methods, qualitative research involves observation of unique events, usually by a single observer and generally without corroboration or replication. Naturally, this over-reliance on a single observer creates serious questions about the reliability of the data obtained. The events reported may have been unique occurrences, or the observer may have unwittingly recorded biased information. Although reliability is an important issue in quantitative data collection, it is usually not defined as problematic by qualitative researchers. They typically argue that just as several artists' paintings of a given mountain may all be "accurate" and yet different, so may different descriptions of a given group vary among observers without being biased or inaccurate.

Aside from identifying the strengths and weaknesses of qualitative research and warning researchers and research consumers about biases and problems inherent in the method, little can be done to maximize the advantages while minimizing the disadvantages. The advantages and the disadvantages are intimately connected. In the process of obtaining a rich, intimate understanding of a people, the researcher is likely to lose some objectivity. On the other hand, too much attention to objectivity will rob the data of its imaginative, impressionistic feelings about how members of the group being studied define themselves and their world.

The strength of qualitative research is its flexibility, but too much flexibility in data collection and organization will produce a useless collection of particulars incapable of being systematically organized and interpreted.

IV. DESIGN OF QUALITATIVE RESEARCH

The important decisions concerning how much the researcher will participate in the group and whether any group members will be told about the researcher's investigative activities are the framework within which a qualitative project is designed. Once the level of participation and secrecy has

been decided, then the specific details about cover and exit stories and the collection of information can be worked out.

Degree of Participation

The qualitative researcher must decide how much he or she is willing to participate in the activities of the group being studied. For example, a researcher studying gang behavior might wish to draw the line at being involved in muggings or attacks on rival groups. Several levels of participation have been identified by researchers who have done **participant observation** (Gold, 1969; Schwartz and Jacobs, 1979; and Spradley, 1980). Gold (1969) suggested four degrees of involvement: complete observer, observer-as-participant, participant-as-observer, and complete participant. It seems to us to be more useful to view participation as a continuum ranging from total participation to nonparticipation. In the cab driving example, Henslin totally participated; he was a taxi driver for a time. In a large community study (Middletown) two of the authors of this text moved their families 1,500 miles, bought a home, enrolled their children in the public schools, shopped in the malls, worshipped in the churches, ate in the restaurants, watched movies in the theaters, and participated in community life for a year (Caplow, Bahr, and Chadwick, 1982, 1983).

In a study reviewed in detail later, Humphreys (1970) functioned as a "Watch Queen" in order to observe homosexual behavior. A Watch Queen does not participate in the homosexual encounter but is present as a lookout who sounds a warning when someone approaches. Thus, Humphreys was only marginally involved in the homosexual behavior but was present and able to observe it.

The nonparticipant observer is an observant bystander, who watches but is not involved in the events being observed. Two of the present authors attended rallies held by Indians in Seattle protesting Bureau of Indian Affairs policies at the time the Trail of Broken Treaties ended in the occupation of BIA offices in Washington, D.C. We did not participate in any of the speech making, singing, or signing of petitions but did attend the rallies and observe the marchers and confrontations with BIA officials in Seattle.

Several factors determine the appropriate level of participation. The qualitative researcher wishes to maximize understanding and awareness of detail, and this encourages participation. Also, some groups are unwilling to admit observers uncommitted to group values, and the only way to obtain accurate data may be to participate or to "pass" as a group member. Both of these factors foster in-depth involvement in the group. On the other hand, some groups engage in activities that, from the researcher's standpoint, may be illegal or immoral. The researcher involved in illegal activity runs the risk of arrest and stigmatization as a criminal;

researchers who participate in what they consider immoral behavior must justify their actions to themselves and significant others in their life and must somehow resolve personal guilt. The researcher must weigh the costs against the benefits of getting close to the data in deciding how far to go in participating.

Degree of Secrecy

A second important decision the researcher must make is how honest he or she will be about the research with the subjects. Degree of secrecy is also a continuum varying from complete secrecy (the spy) to open identification as a researcher. At one end of the continuum, the researcher joins the group pretending to be a convert or an ordinary member. Group members are unaware that they are the subjects of research, and the recording of observations is done covertly. Henslin, in his total participation and total secrecy study, hid a tape recorder under the seat of his cab and taped conversations with his passengers. PFFI, Inc., covertly collected fingerprints, video recordings, and other evidence. Bahr and Chadwick in their nonparticipation and total secrecy study strapped tape recorders to graduate students posing as protestors who marched on the Bureau of Indian Affairs' Seattle office. The microphones were held in a clenched hand, and the wires ran down the sleeves of their coats. Observations about the size of the crowd and its activities were recorded by bringing the hand close to the mouth as though coughing, and then talking into the recorder. Conversations or speeches were recorded by placing the hand as close as possible to the individual speaking. In none of these studies did the participants know that social scientists, or in the PFFI "sting," police officers, were watching and recording their activities.

Under the condition of partial secrecy, the researcher may tell some members of the group about the research. Gatekeepers, or persons who admit members to the group, may be told about the research and asked to allow researchers to join. The leader of a religious group may be asked to allow researchers access to the group. The warden of a prison or the director of a mental hospital might provide the researcher with entrance into the prison or mental hospital, similar to that of any other prisoner or patient. The president of a medical school might admit a researcher to the school, although the researcher lacks the necessary formal qualifications. In return, information identifying problems, information that might help in managing the institution, or data that evaluate specific programs may be promised by the researcher.

The veil of secrecy may be breached even more, and lower-level administrators or leaders may be told of the research so that they can give the researchers better access to the contexts being studied or can protect them from danger. In a prison, the warden may explain the research to

senior-level guards, or perhaps to the entire prison staff, so that the researcher will be allowed to mingle with the prisoners in a variety of situations. Obviously, the more people who know that the participant is a researcher, the greater the likelihood that people try to "manage" the impressions they convey to put themselves and their group in a favorable light.

As more people know they are the subjects of study, the other end of the secrecy continuum is approached, the condition of open, honest disclosure by researchers of their identity, objectives, and methods. A project using such full disclosure practices was one directed by Glock and Bellah (1976), who studied nine new religious movements in the San Francisco Bay area. In all but one of these groups the researchers told the leaders and lay members about their project. The researchers report that their openness produced a corresponding openness in the members of the movements. Glock and Bellah say that they were given access to official records, were allowed to observe rituals, and were able to interview members to a greater degree than would have been possible if they had joined covertly and tried to obtain the data secretly.

Decisions about extent of participation and degree of secrecy are usually established independently, but in some research they are closely linked. Some groups will not allow observation by an outsider. For instance, the qualitative study of gangs who engage in criminal activity typically would require a high degree of participation and total secrecy. A researcher who refused to participate in all gang activities would probably be ridiculed or driven from the group. If his real objectives were known, his physical safety might be threatened.

Nowadays researchers are bound by professional ethics to be honest with their subjects, at least after the data have been collected. Deception is defined by many as both dishonest and exploitative. On the other hand, the benefits of studying behavior in settings inaccessible to researchers sometimes are seen to outweigh the ethical constraints that would dictate honesty at all costs. The greater the potential for research subjects to behave in more socially desirable ways when they know they are being watched, and the greater the perceived threat to the group if its activities were exposed, the more researchers are likely to feel justified in deceiving their subjects.

V. DOING QUALITATIVE RESEARCH

Selection of a Topic

The first step in a qualitative research project is to select a social context or group for study. For some qualitative researchers selection of the target population is the only formal step of advance preparation. The range of

preparation extends from obtaining relevant background information and/or talking to people knowledgeable about the group to the extreme of learning a language and becoming so familiar with a people's activities that the researcher can pass as a member. As for conceptual or theoretic preparation, some qualitative researchers outline general issues that they intend to examine or, in a loose sense of the word, test during their participation. Some go even further, creating systematic protocols or paradigms to guide their observations and questioning.

One of the better known qualitative studies of white prejudice toward blacks is John Howard Griffin's *Black Like Me*. Griffin was haunted by the question of what life was like for a black person living in the South. He asked:

If a white man became a Negro in the Deep South, what adjustments would he have to make? What is it like to experience discrimination based on skin color, something over which one has no control? (Griffin, 1960: 1)

Griffin decided to try to pass as a black person. He discussed the idea with a black friend, the owner of an international black magazine. Additional background information was obtained from discussions with the magazine's editorial director and from three FBI agents in the Dallas office, whom Griffin notified about the study.

Determining the Degree of Participation and Secrecy

The qualitative researcher has to weigh the desire to get close to the data against the dangers involved and has to decide the limits of his or her active participation. The possible consequences to persons being studied and potential costs and benefits to the researcher and the ultimate consumers of the research must be considered and balanced in setting these limits.

Griffin intended to have only casual encounters with both blacks and whites during his study and did not anticipate any dangers to the people he would observe. His friends argued that the project was fraught with potential dangers to himself from prejudiced whites or from angry blacks who might discover his real identity. The reality of these dangers was reinforced by FBI agents. In spite of the warnings, Griffin decided to darken his skin and live for six weeks as a black man in the Deep South. He was convinced that the outcome, his revelation of what it felt like to be black in southern America would improve race relations and, ultimately, the situation of black Americans.

There was the decision about degree of secrecy, whether to infiltrate the black community covertly, to request permission from some black leaders, or to inform everyone about the objectives of the research. Griffin decided that to understand racial prejudice and discrimination he had to

approach both blacks and whites as a black man, without either population knowing who and what he really was. His plans were approved by his wife, and Griffin alerted several friends and the FBI about what he intended to do. In the major cities he visited during the project there usually was someone who knew about the study to whom he could turn if he needed help.

Developing Cover and Exit Stories

Meaningful participation in a social group requires gaining entree, which is often a problem. Wax stresses the importance of motivating the group to accept the researcher.

A human being may immerse himself in a book, a game of chess, and the study of Arabic, or in a bathtub . . . but he cannot, by his own will and determination alone, immerse himself in another *living* group or society. If he manages to squeeze, step, or even dip into a group of living people, it is because the people who are already there invite him or let him in, or at least move over and give him a place to stand. (Wax, 1971: 42-43)

If the participation is conducted in a group setting where everyone knows about the researcher, then cover and exit stories are not necessary. The problem of entrance and exit are more difficult when the participation is covert. Secrecy requires a "natural" entrance and exit, and these may be difficult processes to set in motion. On the other hand, if a powerful leader, such as a prison warden or mental hospital director, agrees to help, then access is readily obtained.

The researcher must examine the group to be studied and learn how new members are recruited, thereby learning possible ways of entree. In some projects, simply hanging around certain neighborhoods, bars, restaurants, or work sites has been sufficient to produce, after a time, entree and acceptance. Some researchers have invented special experiences, credentials, or background characteristics to impress the group members and speed the entry process.

Griffin's major obstacle to acceptance in the black community was his physical appearance. He selected New Orleans as his point of entrance and worked with a prominent dermatologist to darken his skin color. A drug used to darken the white skin spots associated with certain skin diseases was prescribed along with exposure to ultraviolet rays from a sun lamp. The normal time required to darken white spots ranged from six weeks to three months. Griffin was in a hurry to speed up the process so he took heavy doses. The physician monitored the treatment with blood tests to reveal any negative side effects. After four days of treatment Griffin "had a dark undercoating of pigment" which he could accentuate with stain. He shaved his head to hide the lack of curl in his hair. He recorded

how he felt when he stood before a mirror for the first time after the treatment.

In the flood of light against white tile, the face and shoulders of a stranger—a fierce, bald, very dark Negro—glared at me from the glass. He in no way resembled me. (Griffin, 1960: 11)

His first contact with the outside world as a black man was to board a streetcar. He sat in the back with several black people and asked one where he could find a good hotel. The man's response signaled to Griffin that his deception was successful and boosted his confidence in his black identity.

In many qualitative studies an exit story is not necessary. If the milieu being observed is one in which participants regularly enter and drop out, the researcher can decide on the basis of the quantity and depth of his or her field notes when to conclude the data collection. In other, more closed settings a legitimate reason for terminating association with the group is required. One does not drop out of a prison, a mental hospital, or a polygamous marriage without some planning and preparation. For the student of prison behavior, a contrived transfer to another prison would maintain the researcher's cover and provide protection from later reprisals. An extreme, effective exit story is a pretended "accidental death," which completely and permanently removes a researcher from the closed setting.

In Griffin's case an exit story was not necessary. After he had lived as a black man for what he thought was a sufficient time, he removed the stain, stopped taking the medication, dressed in harmony with the standards of white society, and simply disappeared from the black community.

Participation

The goal of most qualitative research is to gain an in-depth understanding of life in a group, community, or society. Ideally, the researcher suspends preconceptions and is immersed in the life of the group. Personal experience is essential—within limits established for self-protection or other reasons. The researcher is involved in the daily activities of the group and alert to potential cues about how the subjects interpret their lives and about the latent as well as the manifest "reasons" for behavior.

Depending upon the framework the researcher eventually selects for organizing the observations, virtually any information has potential value. However, the qualitative researcher is likely to find the following five categories of information especially useful.

Behavior Members of any group or community exhibit certain behaviors and not others. Patterns of behavior (not discrete, unrelated acts) usually are relevant to understanding social life. For example, blacks may

respond to ethnic discrimination in many different ways. Some may complain to civil rights organizations, some may talk to their neighbors, others may drop out of institutions or settings where they are victims of discrimination, and some may ignore the issue entirely. Many other types of behavior also have potential relevance to understanding a community's reaction to ethnic discrimination. Some black people may decide to send their children to college, others will encourage their young people to enter military service. By observing specific behaviors and trying to link them into organized sequences or patterns, the researcher may acquire insights about community life not apparent either to the casual observer or to members of the community.

Meanings One of the strengths of qualitative research is its concern with the meanings people attach to their experiences. A careful observer can learn something about the meanings a group attaches to local traditions or patterns of behavior by noting which actions are recalled with satisfaction, which people are treated with scorn, and which traditions are accorded respect. Information about the meanings associated with certain events, if it is organized with imagination and insight, may reveal much about the real values and priorities of a community, in contrast to the values they profess or recognize themselves.

Problems One useful way to understand people's expectations is to pay close attention to instances where expectations are violated and to observe individual and group reactions to such deviance. Anything that people define as "trouble" deserves special attention, if only because what the community defines as problematic or troublesome establishes the limits of acceptable behavior as well as prompting revealing cases of reinforcement of community norms via the imposition of negative sanctions.

Interactions It is possible to learn a great deal about a group by looking for patterns in interpersonal interaction. Here the focus is upon the shared characteristics that promote acceptance and communication and those that inhibit or prohibit personal interaction. When people interact in some visible division of roles, perhaps one person being a leader, another a lieutenant, and still another a jokester, aspects of the group structure and stratification are laid bare. It is also important to note the characteristics associated with access to leadership and power in the community, such as particular ancestry, minimum educational attainment, possession of wealth, or other factors.

Careful assessment of the different subgroups that make up a community may also be important. The interaction among community leaders and between the leaders and community members may provide some

interesting clues about the status structure and influence patterns that characterize the community.

Discrepancies It is often very fruitful for a researcher to note differences between people's descriptions of behavior and the behavior as observed by the researcher. Similar discrepancies in the accounts of the same event by persons in different subgroups may reveal class-biased interpretations. The researcher should also be aware that although members of the group are in some ways the best possible experts on themselves, they are so close and so intimately involved in their own group that they may fail to note some of the forces that act on them or they may not comprehend some of their reasons for acting, believing, or interpreting events as they do.

Griffin's data collection ended after a harrowing experience on a bus traveling between Tuskegee and Atlanta. During the trip the driver and a white bully tried to force two black young people to give up their seats to whites. In addition, Griffin explained that he was the last to leave the bus and that as he did so, an elderly white man gave him a particularly scornful look. This was the final indignity; Griffin had had enough of the world of black Americans.

It was a little thing, but piled on all the other little things it broke something in me. Suddenly I had had enough. Suddenly I could stomach no more of this degradation—not of myself but of all men who were black like me. . . . In the men's room, I entered one of the cubicles and locked the door. For a time I was safe, isolated; for a time I owned the space around me, though it was scarcely more than that of a coffin. (Griffin, 1960: 140)

Exit

Sometimes ending a qualitative research project consists simply of not showing up at the group's haunts. In other studies disengagement requires planning and tact so that the researcher leaves the group with minimal disruption and without creating feelings of exploitation or anger among subjects. We have stressed the ethical responsibility of the qualitative researcher, noting that he or she must do everything possible to prevent psychological, emotional, social, or physical injury to the persons studied. A sudden awareness that the group has been "taken in" by a researcher's deception may be painful to some. Also, the group may have become dependent upon the researcher in some ways and his or her disappearance may have negative repercussions for the group as a whole as well as for the members the researcher knew best.

In the project that produced *Black Like Me*, following the degradation suffered on the fateful bus ride, Griffin locked himself in a stall in a men's restroom and prepared to leave the black experience.

I took out my cleansing cream and rubbed it on my hands and face to remove the stain. I then removed my shirt and undershirt, rubbed my skin almost raw with the undershirt, and looked into my hand mirror. I could pass for white again. (Griffin, 1960: 140)

He then unobtrusively left the black restroom and reentered the white community.

Later Griffin and a photographer returned to New Orleans and Griffin resumed his Negro appearance to photograph the beginning of his odyssey. After the first articles reporting the study appeared in his friend's magazine, he was invited to appear on regional and national TV talk shows, including the Dave Garroway show and the Mike Wallace show. Shortly thereafter hate phone calls were received by his parents and wife. Griffin was hanged in effigy in his home town of Mansfield, Texas. A half black-half white dummy with a yellow stripe down its back and his name attached to it was hung from a traffic control light in the center of town.

Eventually, his parents sold their home and moved to Mexico. Griffin relocated his wife and children and then waited in Mansfield an extra month to allow the bigots one final opportunity to make good on their threats. After a second deadline for an attack on him passed uneventfully, he joined his family in another area and tried to start a new life.

Griffin's experiences illustrate the dangers that may accompany qualitative research. They also illustrated for him the greater understanding available to the qualitative researcher. Here is his description of his comparison of the utility of the two types of data.

I had spent weeks at work, studying, correlating statistics, going through reports, none of which actually help to reveal the truth of what it is like to be discriminated against. They cancel truth almost more than they reveal it. I decided to throw them away and simply publish what happened to me. (Griffin, 1960: 158)

Analysis of Data and Writing of Report

The qualitative researcher analyzes the data, which generally consist of field notes, to identify significant events, feelings, and patterns of behavior. The quantification of variables and related statistical analysis, if they are included, are only a minor part of qualitative research. The research report is basically a description of what the researcher did, with special emphasis on the researcher's idea of how the subjects interpreted their social world and what their actions meant to them.

The qualitative researcher must be especially careful to protect the research subjects. Names, places, dates, and events must be disguised so that shame, ridicule, persecution, or other social and psychological harm will not befall the people or group whose lives he or she has been permitted to enter and describe.

V. ILLUSTRATIVE EXAMPLES

Festinger, Riecken, and Schachter's "When Prophecy Fails"

Selection of a topic Festinger and his colleagues (1956) had some theoretical notions about how people in religious sects dealt with prophetic failure. The theory had been tested using historical accounts of such incidents, but the historical data were insufficient to test the theory. The researchers hoped to locate a contemporary social movement in which specific prophecies or predictions had been made about events in the near future (Festinger et al., 1956: V).

The following headline appeared in September 1954 in a newspaper the researchers called the *Lake City Herald*:

Prophecy from Planet Clarion Call to City: Flee that Flood. It'll Swamp us on Dec. 21, Outer space tells Suburbanite.

The sect associated with the prophecy seemed made to order for the researcher's needs. A visit to Mrs. Keech, the leader of the "Seekers," confirmed that there was a group of believers that accepted the prophecy that the flood was soon to occur.

Mrs. Keech recounted that one winter morning nearly a year previously she had awakened at dawn with a tingling in her arm. She felt moved to pick up a pencil and a pad and begin to write in a handwriting different from her own. She developed an ability to receive messages through her writing and eventually established a relationship with beings called the "Guardians" from the Planet Clarion who informed her of an impending flood. The first message announcing the impending disaster had been received in early August.

The earthling will awaken to the great Casting. Conditions to be fulfilled of the lake seething and the great destruction of the tall buildings of the local city—the cast that the lake bed is sinking to the degree that it will be as a great scoop of wind from the bottom of the lake throughout the countryside. You shall tell the world that this is to be, for such it is given. To you the date only is secret for the panic of men knows no bounds. (Festinger et al., 1956: 55)

Another message received ten days later made it clear that the predicted upheaval was to extend way beyond the midwestern state in which Mrs. Keech lived.

This is not limited to the local area, for the cast of the country of the U.S.A. is that it is to break in twain. In the area of the Mississippi, in the regions of the Canada, Great Lakes and the Mississippi, and the Gulf of Mexico, into the Central America will be as changed. The great tilting of the land of the

U.S. to the East will throw up mountains along the Central States, along the Great New Sea, along North and South—to the South. The new mountain range will be called the Argone Range which will signify the ones who have been there are gone—the old has gone past—the new is. This will be as a monument to the old races; to the new will be the Altar of the Rockies and the Alleghenies. (Festinger et al., 1956: 56–57)

Mrs. Keech's prediction of a specific event on a given day and the gathering of those who believed her provided Festinger and his associates the opportunity to test their ideas.

Determining the degree of participation and secrecy The researchers felt that complete participation was necessary to obtain the information necessary to test their theory. They wanted to be actively involved in all the Seeker meetings, to be present during informal gatherings, to watch group members, and to get to know them well so that their feelings about events could be subtly probed. The initial contact with Mrs. Keech convinced the researchers that nonbeliever observers would not be welcome for any extended period, and so the decision was made to participate fully.

Festinger and his staff tried to keep the observers' influence on the group to an absolute minimum. "We tried to be nondirective, sympathetic listeners, passive participants who were inquisitive and eager to learn what others might want to tell us" (Festinger et al., 1956: 237). However, as we will see, noninfluence was impossible to achieve.

The Seekers' antipathy toward nonbelievers also convinced the researchers that secrecy was essential. Thus the study was conducted by participant observers without the knowledge or the consent of the Seekers.

Cover story and entry The Seekers had two gathering places, Lake City (Mrs. Keech's home) and Collegeville (Dr. Armstrong's home). A man and a woman were assigned to each site. In Lake City the woman observer called on Mrs. Keech and said that she had attended a religious and philosophical meeting and that the man next to her had told her about the Seekers. The observer acted embarrassed and stated that she had come to the house on an impulse. Mrs. Keech took the bait, invited her into the house, and told her about the Seekers and the anticipated flood. The observer received permission to return. The other Lake City observer told Mrs. Keech that he had read about her in the newspaper and that he wanted to learn more about the movement. She invited him to Seeker meetings.

In Collegeville more creative cover stories were invented to gain immediate and complete access. The male observer attended open meetings of the "elementary" Seekers and tried to establish a relationship with Dr. Armstrong, a physician at the University Health Center, who was a convert to the Seekers. The observer had little success in arousing Dr. Armstrong's interest and so the researchers decided to have him fabricate a supernatural

experience (Festinger et al., 1956: 239). The observer told Dr. Armstrong that he and a companion had had a strange experience while driving in Mexico. According to the story, the observer and his companion had picked up a hitchhiker, an elderly woman, who had sat in the back seat warning them about impending disasters. She eventually fell silent and they assumed she was asleep. Later they turned to ask where she wanted to be dropped off and discovered she had vanished. They said they had been driving at a high speed and had not heard the car door open or any noise suggesting that she had jumped or fallen from the car. The story interested Dr. Armstrong and the observer was invited to attend the meetings of the Seekers in Armstrong's home. The female observer in Collegeville was also armed with a psychic experience. She went directly to the Armstrong home and told them of a dream she had had.

I was standing on the side of a hill. It wasn't a mountain, yet it wasn't exactly a hill: and I looked up and there was a man standing on top of the hill with a light all around him. There were torrents of water, raging water all around, and the man reached down and lifted me up, up out of the water. I felt safe. (Festinger et al., 1956: 240)

Not surprisingly, the Armstrongs interpreted the dream to mean that the Guardians had guided this young woman to them in order that she might be saved from the impending disaster.

Thus, considerable deception was employed by the researchers to gain access to the Seekers. We will see later how this deception eventually had a significant impact upon the group.

Participation All the observers were in place by November 19, about a month before the predicted disaster. Observation was maintained for six weeks, including two weeks beyond the crisis of the failed prophecy. Participation was extremely tiring. The observers had to spend almost all their waking hours with the Seekers, and meetings frequently lasted until late at night. Personal affairs, including family, friends, work, and school had to be neglected.

The observers were instructed to learn as much as possible about each member of the Seekers. They tried to judge the strength of the members' belief in the ideology, their commitment to the group, the actions they had taken in preparation for the flood, and finally, the degree to which they tried to convert others to the cause. The data collected by the observers consisted of "anecdotal accounts of events that took place in their presence, reports to them of actions that members had taken earlier or elsewhere; factual or attitudinal data elicited in interviews or conversations with members; and the content of talks or assertions made to the group as a whole" (Festinger et al., 1956: 250).

The observers could not take notes openly, and they had to find opportunities to leave the group to record their observations. Occasional trips to the bathroom, slipping out on the back porch, or taking a walk to get a breath of fresh air were the main ploys used. Each night when they returned to their homes, the observers dictated an account of the day's activities, fleshing out the brief jottings made during the day.

The observers tried to minimize their impact on the group but were unable to prevent themselves from becoming part of the group and influencing its actions. The most serious impact occurred when Mrs. Keech commanded the man who had told the story of the Mexican woman's disappearance to lead a meeting. Not knowing what to do, he stalled for time by suggesting that the group silently meditate. During the painful silence that followed, one of the group suddenly began to moan and breathe very deeply. She then began to gasp, "I got the words" over and over again. With her eyes closed, and shaking with emotion, she delivered a message from the Guardians. From this time on, this type of medium experience was an integral part of the Seekers' activity. Thus, inadvertently, the observer introduced a new and rather dramatic ritual.

Also, the fact that the four observers had joined the group and that two of them reported psychic experiences reinforced the faith of other believers in the Guardians and the approaching flood. The arrival of the four converts was interpreted as evidence that the Guardians were gathering the elect to save them from the coming disaster.

The two observers in Lake City were pressured by other Seekers to quit their jobs and make arrangements for the coming disaster on December 21. Their attempts to maintain neutrality had a major impact on other group members.

One observer persistently avoided making any statement about his plans; the other waited until the 17th, and then announced that her job had been terminated. Yet their evasion of these requests and their failure to quit their jobs at once were not only embarrassing to them, and threatening to their rapport within the group, but also may have had the effect of making the members who had quit their jobs less sure they had done the right thing. In short, as members, the observers could not be neutral—any action had consequence. (Festinger et al., 1956: 244)

As December 21 approached, the Guardians sent instructions to the Seekers to prepare to be picked up by a flying saucer so that they would escape the predicted flood. The Lake City morning newspaper on December 17 contained some rather antagonistic articles about the Seekers and during the morning a phone call was received from a man identifying himself as "Captain Video." This man informed the group that they would be met by a flying saucer at 4 P.M. that afternoon. Although there was suspicion that the phone call was a hoax, it was decided that the group

could not chance ignoring it. The Guardians had announced that one condition of entrance into a flying saucer was that the person have no metal on them. Rings, watches, eyeglasses with metal frames, belt buckles, snaps, metal buttons, zippers, eyelets in shoes, and fasteners were removed. This meant that the men ripped the fasteners out of their pants and removed their belts. The women removed their bras and any metal buttons or fasteners from their clothing. Most wore heavy socks or slippers in lieu of shoes, which have nails holding on the heels.

A TV camera waited in front of Mrs. Keech's home as the seekers huddled in the backyard while the 4 P.M. deadline quietly passed. Later that evening, around midnight, Mrs. Keech received a message that a flying saucer was actually on its way. The haste to move to the yard is illustrated in the observer's notes.

We went out in the backyard. It was cold and snowing and the ground was very wet. Marion, Daisy, Thomas, and Edna were all there. Marion told me about the message, and asked me again about metal. She wanted to know about my shoes, and Edna said I would have to take my shoes off because they had nails in them.

Mark came back in the house with me. I took my shoes off and Mark started to rip the heels from the shoes. I stopped him and said, "Don't do that. Just get me a couple of pairs of wool socks and some bedroom slippers." He did this and then pointed out that the buttons on my suit did have metal on them. I ripped the buttons off my coat.

We got back outside and Edna took me aside and said, "How about your brassiere? It has metal clasps, doesn't it?" I went back in the house and took my brassiere off. The only metal on me was the fillings in my teeth and I was afraid someone would mention those. (Festinger et al., 1956: 157)

The flying saucer did not keep the appointment, and at 3:30 A.M. a very tired and cold group of Seekers returned to the house. The failure of the prophecy of the flying saucer did not overly discourage the group, as it was eventually interpreted as a "drill" for the real evacuation to come.

The tension-packed days passed slowly as December 21 and the prophesied flood approached, and the Seekers waited for new instructions. At 10:00 in the morning of December 20, the anxiously awaited news arrived.

At the end of midnight, you shall be put into parked cars and taken place where ye shall be put aboard a parch (flying saucer) and ye shall be purposed by the time you are there. At that time you shall have the fortunated ones forget the few who have not come—and at no time are they to be called for, they are but enacting a scene and not a person who should be there will fail to be there and at the time you are to say "What is your question" . . . and at no time are you to ask "What is what" and not a plan shall go astray and for the time being be glad and be fortunated to be among the favored. And be ready for further instructions. (Festinger et al., 1956: 158–59)

Shortly before midnight a message was received informing the group to get their overcoats and stand ready for evacuation. Again the deadline passed but no flying saucer appeared. Later another message was received that there had been a slight delay. The group waited patiently, but again the saucer did not materialize. Eventually, the group was forced to recognize not only that the flying saucer was not coming, but that there would be no flood. The individual Seekers, understandably, had trouble accepting the disconfirmation of the prophecies. Later Mrs. Keech received another message that explained why the predictions had not been fulfilled. The Seekers were commended for their faith and told that they had spread "light" and that this light was the flood that would sweep the world. In other words, the destruction had been called off because of the Seekers' faithfulness, and thus evacuation was not necessary.

Exit It was easy for the observers to exit the Seekers as the group disintegrated a few weeks after the prophecy failed. The police threatened to serve a warrant on Mrs. Keech because of the disturbances she had caused and she went into hiding, later moving away using an alias. The Armstrongs sold their home in Collegeville and moved elsewhere. The observers were free to resume their former lives with no worry about injury or reprisal from members of the Seekers.

Analysis and report The 65 hours of tape-recorded field notes were transcribed and analyzed by the research team. The qualitative nature of the data and the justification for excluding quantitative techniques were made explicit:

Our material is largely qualitative rather than quantitative, and even simple tabulations of what we observed would be difficult. Owing to the complete novelty and unpredictability of the movement, as well as the pressure of time, we could not develop standard categories of events, actions, statements, feelings, and the like, and certainly could not subject the members of the group to any standardized interview, in order to compare indices before and after disconfirmation. (Festinger et al., 1956: 252)

The emergence of the Seekers and the prophecies and their subsequent disconfirmation allowed the researchers to test their theoretical ideas which later became widely known as *cognitive dissonance theory*. Certain key theoretical notions such as the group's zealous proselytizing immediately following the disconfirmation were supported. This behavior was interpreted as a means of resolving the dissonance associated with the failure. The research report, *When Prophecy Fails*, is an exciting and insightful account which has had a significant impact in social psychology.

Humphreys' "Tearoom Trade"

Selection of a topic Humphreys' initial interest in homosexual behavior occurred during a ten-year period when, as an Episcopalian minister, he counseled homosexuals. Later, during graduate study in the Sociology Department of Washington University at St. Louis, he wrote a research paper on the topic. His graduate advisor pointed out that very little was known about male homosexuality and encouraged Humphreys to consider doing a dissertation on this subject. Humphreys' statement of the research problem was that he intended to describe the functioning of the homosexual community.

My concern in this study has been with the description of a specific style of deviant behavior and of the population who engage in that activity. Beyond such systematic, descriptive analyses, I have tried to offer, in the light of deviance theory, some explanation as to why, and how these people participate in the particular form of behavior described. I have not attempted to test any prestated hypothesis. (Humphreys, 1970: 22)

Determining the degree of participation and secrecy Humphreys visited gay bars and public restrooms identified as sites of homosexual activity and talked to homosexual men in an effort to find a workable degree of participation that would provide him with necessary data but not compromise his personal values. His own values mitigated against full participation, but at the same time nonhomosexuals were excluded from the action. Humphreys learned that a man who remained in a public washroom longer than five minutes was automatically identified as a homosexual on the make or a member of the vice squad. The solution was to play the role of "Watch Queen" (lookout). During his initial observations, he learned that there are three types of watch queens. "Waiters" are those men who look out for others while waiting their turn to be involved. "Masturbaters" are men who serve as lookouts while they also engage in masturbation. "Voyeurs" are Watch Queens who derive their pleasure from watching others engage in homosexual behavior. This last role was ideal for Humphreys' purposes and was assumed by him. He was able to observe all the behavior without alarming the participants or disturbing the action.

Discussion with members of the gay community convinced Humphreys that the study would have to be conducted in total secrecy. During the 1960s when the study was conducted, homosexuals were subject to arrest and prosecution, public ridicule, and loss of employment, all of which made them vulnerable to blackmail. To make matters worse, a gay informant warned Humphreys that homosexuals in the St. Louis community were particularly wary of sociologists (Humphreys, 1970: 242). Humphreys was told that a graduate student at another university had failed to disguise the names of gay respondents in research for a Master's thesis. Therefore,

Humphreys decided to do the study under total secrecy. However, as the study progressed over a two-year period, he took into his confidence twelve members of the homosexual community whom he called the "Intensive Twelve." Having established good rapport with these twelve men, he explained what he was doing and obtained their cooperation in submitting to several lengthy interviews about their backgrounds and their feelings, beliefs, and practices relating to homosexuality.

Cover story and entry Humphreys passed himself off as just another "gay guy" as he made the rounds of ten gay bars in the St. Louis metropolitan area. Gradually he became known and accepted and was able to attend private parties and the annual "drag ball." His presence in places of instant sex, such as public restrooms, a local bathhouse, and certain movie theaters, was eventually accepted. However, his gradual entrance into the homosexual world took several months to accomplish.

Participation Participation involved driving to a particular restroom at a time homosexual encounters were likely to occur. Humphreys would enter the restroom, station himself by one of the windows, and by his actions, make it clear he was willing to serve as Watch Queen. Most visits lasted 10 to 30 minutes. He made an attempt to rotate the restrooms visited and the time of his visits, so as to observe as broad a segment of the homosexual community as possible. Using these techniques, he observed 120 different homosexual encounters which took place in 19 different restrooms located in the city's public parks.

Although Humphrey's research methods were largely qualitative, he also sought quantitative data where possible. A codesheet-diagram of the typical restroom was used to record the location and activities of each of the participants. Immediately following an observation session, Humphreys would sit in his car and fill out the codesheet. In addition, a tape recorder was hidden under a pasteboard box on the front seat of the car, into which he dictated field notes.

Because homosexuals were severely stigmatized and subject to legal prosecution in the 1960s, most were unwilling to talk about themselves and their backgrounds. Therefore, Humphreys had to find other ways to learn about the subjects' characteristics. After observing a man engage in a homosexual act, Humphreys would follow the man out of the restroom and note the license plate number of his car. As the observation portion of the study continued for a year, he had time and opportunity to check and double-check the plate numbers of the 134 men he had selected. State motor vehicle registration offices cooperated and Humphreys gained access to the automobile registration of 100 men. The other 34 were unavailable because of a variety of reasons, mainly that the car had been recently sold. The records included the men's name, address, marital status, as well as

information about the age, make, and model of their automobile. Humphreys now possessed extremely sensitive information: the names and addresses of 100 men whom he personally had watched perform homosexual acts.

Humphreys used the auto registration information as a step to more intensive data collection. Armed with the addresses, he located the men's homes and walked their neighborhoods, thus learning about their social class background and lifestyle. Particular attention was given to objects visible in the yards: swing sets, tricycles, and other toys were evidence of children; a St. Mary's shrine implied that the man belonged to a Roman Catholic family; and boats, campers, trailers, and so on suggested that the family had an outdoor lifestyle.

In spite of neighborhood observation data, Humphreys did not have sufficient detail about the personal characteristics of the men themselves. He knew that he could not approach the men in their homes and tell them that he had watched them have homosexual sex in a public restroom and casually mention that he would like to ask them a few questions. In Humphreys' words (1970: 41): "I had no desire to conclude my research with a series of beatings."

He finally managed to obtain additional data on the men by including them in the sample of a large social health survey being conducted in the community. The interview elicited information about family background, socioeconomic characteristics, personal health, religious participation, employment, political attitudes, friendship networks, and information on the marital relationship including sex. Humphreys did most of the interviewing of these men himself, although a fellow graduate student did assist some. Humphreys altered his appearance by changing his hair color, donning glasses, wearing different clothes, and driving a different automobile. Although a year had elapsed between the observations in the public restrooms and the interviews, he reported some anxiety as he knocked on the doors of the men's homes. Eventually, interviews were obtained from 50 of the men and that 50 percent response rate was similar to that in the general health survey. A matched control group from the health survey sample was used as a comparison sample for making systematic contrasts between known homosexuals and presumed nonhomosexual males.

During the two years of the study Humphreys amassed a remarkable data set. He had observed 134 men engage in 120 homosexual actions, had obtained considerable background information on 100 of these men, and had conducted detailed interviews with 50 of them.

Humphreys' experiences while practicing participant observation reveal some of the dangers of this research method. One summer afternoon he was standing about ten feet outside a public restroom when a car at a high speed approached up the walk-pathway to the restroom. Two men that Humphreys guessed were policemen jumped out of the car and ran

into the restroom, leaving the car doors open and the motor running. A few minutes later another man emerged, replacing what Humphreys was convinced was a wallet, implying that the vice squad men had elicited a bribe. The two police officers came out of the restrooms, approached Humphreys, and demanded his name and address. He refused and was arrested.

A little over two hours later I was released from police custody. During that time in the precinct station, I was questioned, frisked, stripped of my wallet and keys, booked on a charge of loitering, locked up in an all-metal cell, and deprived of cigarettes. In spite of much pleading on my part, I was never allowed to make a phone call. They insisted upon calling my wife for me. Suppressing her laughter, she phoned my attorney. After an extensive lecture about giving officers "my name, rank, and serial number" and about the dangers of "hanging around those park restrooms," I was released on summons. Because I am a minister, and I have an astute attorney, my case never appeared in court. I am an arrest statistic not a conviction statistic. (Humphreys, 1970: 95-96)

The police were not the only danger. A month later Humphreys and four other men in a restroom were attacked by a gang of youths who yelled insults at the men inside. As the yelling escalated, one of the boys inserted a stick in the hasp on the outside of the door, locking the men in. For half an hour the men "endured a barrage of stones and bottles, which broke every window in the facility, scattering glass the full width of the floor" (Humphreys, 1970: 99). Eventually the youths tired of the game and left and the men escaped.

Humphreys continued the project for two years, April 1967 to April 1969, before reaching the point where he felt that he had seen and experienced enough. In spite of the many precautions he had taken, he had worked under considerable risk of detection and possible injury to himself, and his data set was a possible source for ruining the careers and personal lives of the subjects of his study.

Exit Exit was easy. He stopped frequenting the public restrooms and faded from the homosexual scene.

Analysis and report The analysis involved reviewing the codesheets and taped field notes and looking for patterns in the behavior. The social background, economic class, and the detailed information in the interviews were analyzed to identify characteristics associated with homosexuality.

The report, *Tearoom Trade*, is a fascinating inside view of the world of casual sex between men. The book was awarded the C. Wright Mills Award by the Society for the Study of Social Problems for being the best book published in 1970 on a critical social issue.

Although the book was applauded by many when it first appeared, Humphreys was also strongly criticized for invading the privacy of his subjects and for collecting sensitive data which might be used to blackmail them or which could result in legal prosecution. The furor caused by the initial publication of this study prompted Humphreys to burn his tapes and materials and to make certain that any identifying passages in manuscripts were deleted. He mentally resolved that if called into court, he would plead the Fifth Amendment and risk a contempt citation rather than reveal the identity of any of the men.

The Chancellor of Washington University felt that Humphreys had committed numerous felonies in the process of conducting the research, and terminated his teaching contract and participation in other research projects. The Chancellor unsuccessfully tried to have Humphreys' Ph.D. degree revoked. Humphreys was also physically assaulted by a faculty member who was angered by the research and Humphreys' defense of it. As Humphreys (1970: 23) has reevaluated his research methods, he has admitted that tracking the men's identity through license plates and then interviewing them under false pretenses was a serious breach of their rights.

VII. SUMMARY

Qualitative research refers to several different means of data collection, including participant observation, ethnomethodology, and ethnographical research. It emphasizes getting close to the data and is based on the concept that experience is the best way to understand social behavior. The testing of theories via the collecting of quantitative data is seen as distorting social reality because counting and measuring force data into irrelevant theoretical pigeonholes. Qualitative researchers suggest that a more accurate understanding of social behavior is accomplished by approaching the setting with as few theories and measurement tools as possible and then immersing oneself into the behavior being studied. The researcher must suspend personal values and feelings and experience the world from the viewpoint of those being studied.

One of the advantages of qualitative research is that behavior is observed in its natural setting. In addition, the qualitative researcher gains an in-depth understanding by being intimately involved with the research subjects. Because preconceived theories and measurement techniques are absent, qualitative techniques allow flexibility to change the direction of the study as the researcher becomes more aware of what he or she is studying.

Qualitative research has a strong potential to violate the rights of subjects, as potentially harmful information may be collected about indi-

viduals without their knowledge or consent. Another problem is that the researcher's participation may change the group. Ideas are given, activities supported, and friendships established that may alter the group so that what the researcher started to study no longer exists. Participant observation of some groups may lead the researcher to violate the law or his or her own moral values. Finally the qualitative researcher runs the risk of "going native," accepting the values of the group being studied and emerging with very biased perceptions.

The researcher's participation may vary from total involvement to merely observing the behavior of the members. A related issue is the degree of secrecy surrounding the researcher's identity and intentions. Complete secrecy is sometimes necessary for admittance into certain groups, whereas in other cases the leaders' cooperation can be obtained.

Qualitative researchers sometimes need a cover or entrance story to gain access to the group they desire to investigate. Sometimes an exit story is helpful for the researcher to disengage from the group. Once admitted to the group, the researcher participates with them until he or she has sufficient experience to describe the behavior in question from the point of view of the subjects.

REFERENCES

- ALFRED, RANDALL H., 1976. "The Church of Satan" in Charles Glock and Robert N. Bellah (eds.), *The New Religious Consciousness*. Los Angeles: University of California Press.
- BECKER, HOWARD S., and BLANCHE GEER, 1957. "Participant Observation and Interviewing: A Comparison." *Human Organization* 16: 28-32.
- CAPLOW, THEODORE, HOWARD M. BAHR, and BRUCE A. CHADWICK, 1982. *Middletown Families*. Minneapolis: University of Minnesota Press.
- 1983. *All Faithful People*. Minneapolis: University of Minnesota Press.
- DEUTSCHER, IRWIN, 1973. *What We Say/What We Do*. Glenview, Ill.: Scott, Foresman.
- FESTINGER, LEON, HENRY W. RIECKEN, and STANLEY SCHACTER, 1956. *When Prophecy Fails*. Minneapolis: University of Minnesota Press.
- FILSTEAD, WILLIAM J. (ed.), 1970. *Qualitative Methodology*. Chicago: Markham.
- FREUDENBURG, WILLIAM and STAN L. ALBRECHT, 1982. "Participant Observation Analysis" in S.H. Murdock, J.K. Thomas, and D.E. Albrecht, *Handbook for Assessing the Social and Special Effects of Nuclear Repository Siting*. A report to the office of Nuclear Waste Isolation. Battelle-Columbus Dept. of Energy, pp. 103-43.
- GLOCK, CHARLES, and ROBERT N. BELLAH (eds.), 1976. *The New Religious Consciousness*. Los Angeles: University of California Press.
- GOLD, RAYMOND L., 1969. "Roles in Sociological Field Observation" in

- George J. McCall and J. L. Simmons (eds.), *Issues in Participant Observation*. Reading, Mass.: Addison-Wesley.
- GRIFFIN, JOHN HOWARD, 1960. *Black Like Me*. Boston: Houghton Mifflin.
- HENSLIN, JAMES M., 1967. "Craps and Magic." *American Journal of Sociology* 73: 316-330.
- 1968. "Trust and the Cab Driver" in Marcello Truzzi (ed.), *Sociology of Everyday Life*. Englewood Cliffs, N.J.: Prentice-Hall.
- 1972. "What Makes for Trust" in James Henslin (ed.), *Down to Earth Sociology*. New York: The Free Press.
- 1974. "Sex and Cabbies" in James Henslin (ed.), *The Sociology of Sex*. New York: Appleton-Century-Crofts.
- HUMPHREYS, LAUD, 1970. *Tearoom Trade*. Chicago: Aldine.
- JACOBS, GLEN (ed.), 1970. *The Participant Observer*. New York: George Braziller.
- MALINOWSKI, BRONISLAW, 1922. *Argonauts of the Western Pacific*. New York: E. P. Dutton.
- SCHWARTZ, HOWARD, and JERRY JACOBS, 1979. *Qualitative Sociology*. New York: The Free Press.
- SPRADLEY, JAMES P., 1980. *Participant Observation*. New York: Holt, Rinehart & Winston.
- WAX, ROSALIE H., 1971. *Doing Fieldwork: Warnings and Advice*. Chicago: University of Chicago Press.
- WHYTE, WILLIAM FOOTE, 1955. *Street Corner Society: The Social Structure of an Italian Slum*. Chicago: University of Chicago Press.

Content Analysis

nine

I. Introduction

II. Strengths and Weaknesses of Content Analysis

III. Sources of Material

IV. Computer Assisted Content Analysis

V. Doing Content Analysis

Statement of the problem

Selection of communications

Operational definitions

*Training coders and checking
reliability*

*Analysis of data and writing of
report*

VI. Exemplary Content Analysis Studies

Violence profile

selection of communications

operational definitions

*training coders and checking
reliability*

*analysis of data and writing
of report*

The emergence of Indian activism

selection of communications

operational definitions

*training coders and checking
reliability*

*analysis of data and writing
of report*

VII. Summary

I. INTRODUCTION

Content analysis is the study of communications to describe social behavior or to test hypotheses about it. A more formal definition stresses the objectivity and systematic procedure that distinguishes content analysis from other analyses of communication:

Content analysis is any technique for making inferences by objectively and systematically identifying specified characteristics of messages. (Holsti, 1969: 14)

In other words, content analysis involves systematically coding messages, or information in them, into categories, thus allowing quantitative analysis. The entire range of human communication, from gestures to textbooks, from billboards to television commercials, is suitable for content analysis.

Studies of the content of television shows are common nowadays. One such study (Silverman, Sprafkin, and Rubinstein, 1979) assessed the extent of explicit sexual behavior on prime-time TV. The universe of content included all programs broadcast by the three major networks, ABC, CBS, and NBC in the New York City area during a week in October 1977. Drama, crime, adventure, situation comedy, variety, specials, and movie programs telecast between 8 and 11 P.M. were videotaped. Seven observers were

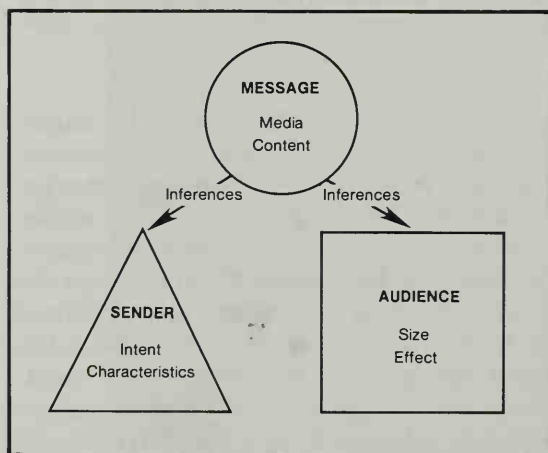
trained to code the material into the following eight behavioral categories: kissing, hugging, ritual touching, supportive touching, professional touching, affectionate touching, suggestiveness, and sexual intercourse. In addition, whether the behavior was physical, verbal, or implied was coded along with the gender and race of the participants. Two coders, one male and one female, independently coded each program. The reliability between coders was reasonable: a correlation of .76 for physical presentations, .75 for verbal presentations, and .55 for implied behavior (a correlation of 1.0 would signify perfect agreement). The lower level of reliability for implied behavior reflects the subjective nature of interpretations about what is implied in human relationships. The results revealed little explicit sexual behavior on television in 1977, but the programming did contain considerable flirtatious behavior and verbal innuendo about sexual activities.

Communications are typically described as having three major components: sender, message, and audience, as shown in Illustration 9.1 (Holsti, 1969; Carney, 1972). The message is available for analysis in terms of explicit themes, relative emphasis on various topics, amount of space or time devoted to topics of interest, and numerous other quantitative dimensions, and reasonable conclusions can be made about the information it contains.

Occasionally messages are analyzed for information about the sender of the communication. The linkage between message content and attributes of sender is often a tenuous one, but some characteristics of the sender are discernible, especially if numerous examples of his or her communication are available.

Researchers also use content analysis to assess a message's effects on an audience. The Pornography and Television Violence Commissions tried,

ILLUSTRATION 9.1 Areas of content analysis.



especially at first, to assess the impact of sexual material or violence in television and movies on those who watched such materials (Commission on Obscenity and Pornography, 1970; Comstock and Rubinstein, 1972). However, making inferences about either the characteristics of the sender or the effects of the message on the receivers is tricky, and sometimes more of an art than a science. At the very least, such inferences about senders or receivers from the assessment of message content must be seen as exploratory rather than definitive.

Content analysis is most often used to describe events or processes in society. Sanders (1974) argues that content analysis is an insightful means to study social change, for the writings of a people—the printed communications as well as their private ones—reflect changes in values, beliefs, and behaviors. For example, one index of discrimination against minority groups is measured by counting the frequency of negative ethnic stereotypes in history books, cartoons in magazines, TV programming, and commercials in magazines, newspapers, radio, or television. Public attitudes toward important issues like ERA, military incursions in adjoining countries, inflation, unemployment, and so on can be assessed by content analysis of letters to the editor or editorials in newspapers. However, it must be remembered in such work that generalization is possible only to the subsample selectively represented, that is, people who write to newspapers and whose letters are printed in the first instance, and newspaper editors themselves in the second. Statements about values held by segments of society can also be inferred from the lyrics of popular songs, although, again, technically speaking, the lyrics represent the songwriter's rather than the listeners' sentiments.

Although content analysis is generally used to describe, it can also be used to test hypotheses. The most frequent type of hypothesis tested is whether values or behaviors have changed over time. The hypothesis that ethnic discrimination has decreased in society, or at least in the society that creates television commercials, could be tested through content analysis of negative stereotypes in TV commercials over a 20-year period.

Content analysis is sometimes useful as a supplement to quantitative analysis in the interpretation of open-ended items in questionnaires or interviews. For example, in a study of religious values in Middletown, U.S.A., the responses to three open-ended questions obtained in interviews with samples of married women in Middletown in 1924 and 1978 were content-analyzed (Caplow, Bahr, and Chadwick, 1983). The three questions were:

1. What are the thoughts and plans that give you the courage to go on when thoroughly discouraged?
2. How often have you thought of Heaven during the past month in this connection?

3. What difference would it make in daily life if you became convinced that there is no loving God caring for you?

Content analysis revealed, contrary to popular beliefs about the decline of religion, a persistence, if not an increase, in the acceptance of religious values. In 1924 (Lynd and Lynd, 1929), 13 percent of the working-class wives rejected the question about there not being a loving God as "un-thinkable," as did 17 percent of the wives in 1978. Fifty-eight percent of the 1924 women and 51 percent of the 1978 respondents reported that life would be intolerable for them or utterly changed if they became convinced that there were no loving God. Significant distress at such a condition appeared in the answers of an additional 18 percent of the wives in 1924 and 23 percent in 1978. Eleven percent of the wives in 1924 and 9 percent in 1978 said that such a conviction would make no particular difference to them. The responses of the women in 1924 and 1978 were so similar that coders could not tell them apart. The words, ideas, and expressions describing religious feelings were nearly identical. The results clearly refute, at least in Middletown, the notion that religion is dying out in American society. The content analysis made a significant contribution to the Middletown study of religion, although most of the information used to identify continuities and changes in religiosity in Middletown was quantitative data from interviews and questionnaires.

Verbal comments recorded during experiments can be content-analyzed to describe the frequency of particular behaviors under certain conditions, or to test a relationship between variables. Chadwick and Day (1972) conducted an experiment testing the relationship between frustration and instrumental aggression. The verbal comments of the subject and a confederate frustrating the subject's attempts to win money during a ten-hour experiment were tape-recorded, then transcribed and analyzed for patterns. Comments were coded into five categories: suggestions, evaluations, pressure tactics, indirect aggression, and aggression. The results revealed that subjects had a "response hierarchy" that they cycled through in dealing with the frustrating confederate, that is, when suggestions or pointed evaluations failed to induce the desired changes in someone's behavior, a more obtrusive approach—pressure tactics or indirect aggression—was apt to be tried (Chadwick and Day, 1972).

Content analysis is also very useful in the analysis of **projective personality tests** such as the Rorschach, the Thematic Apperception Test, and the Twenty Statement Test. Projective techniques supposedly measure a respondent's unconscious or repressed perceptions or feelings by having him assign a meaning to abstract stimuli (like inkblots), tell a story, draw a picture, complete a sentence, or role-play. The objects identified, stories told, pictures drawn, sentences completed, or roles played are then content-analyzed for patterns characteristic of certain personality traits.

For example, Attkisson and his associates (1969) wanted to know if being religious affects the value patterns apparent in responses to the Draw-A-Man Test. Over 200 Baptist, Episcopal, Presbyterian, Church of Christ, and Roman Catholic theology students, religiously involved Methodist undergraduates, and Roman Catholic priests were given the Draw-A-Man test. Subjects each were supplied with two sheets of unlined paper and a pencil and instructed:

On the first sheet of paper, I would like you to draw a person. Be sure that you draw a whole person and not just a head. We are not interested in how good an artist you are. This is not a test of artistic ability. (Attkisson, Nandler, and Shrader, 1969: 28-29)

When the first drawing was completed, the subjects were instructed to turn it over and then on the second sheet of paper to draw a person opposite in sex from that in their first drawing.

The drawings were analyzed by three independent coders for religious symbols, persons, activities, or places. Religious symbols included figures of Christ, persons holding Bibles, a nun, and some elaborate religious symbols. A wide variety of nonreligious figures were drawn, including farmers, sailors, golfers, football players, basketball players, supermen, superwomen, soldiers, women in bathing suits, women in low-cut dresses, spacemen, and cowboys. The assignment of figures to the appropriate category was apparently not difficult. It turned out that the Draw-A-Man Test was not a sensitive measure of religious values, for only eight out of 400 drawings had a religious content. Although in this case the projective test proved not to be a useful technique for assessing religious values, the project does illustrate how content analysis can be used to validate and interpret projective tests.

Recent criticism has challenged the use of projective techniques because clinicians lack the necessary skills in content analysis to adequately score and interpret them. It is argued that content analysis of the Rorschach may be the best means of interpretation (Howes, 1981: 346) and that the Rorschach has considerable unrealized potential that might be achieved if formal content analysis were used in assessing the test results (Aronow and Reznikoff, 1976).

II. STRENGTHS AND WEAKNESSES OF CONTENT ANALYSIS

One important advantage of content analysis is that it does not use up any of that scarce commodity, human research subjects. Content analysis is usually nonreactive in that no one is interviewed, no one has to fill out a questionnaire, no one must come to a laboratory. Moreover, content anal-

ysis tends to be relatively inexpensive. The materials for content analysis are usually readily available. Libraries have archives of public communications (newspapers, magazines) and often of private communications (letters, diaries), at least of certain historical figures. Video and audio recording equipment, cameras, and copiers which greatly facilitate content analysis can be expensive but often need not be purchased. Coders' time is expensive, but not when compared with the costs of survey or experimental research.

Another important advantage of content analysis is that it can be used when the researcher is prevented from surveying or observing the population being studied. For example, Chai (1977-1978) studied the political conflict in Red China following the 1976 death of Mao Tse-Tung by content analysis of official obituaries. It was impossible for American scholars to survey or to observe firsthand the Chinese reaction to Mao's death. However, forty obituary notices, which were sent to the Central Committee of the Communist Party of China by eleven military regents, three cities, and twenty-six provinces, were available and were systematically analyzed. Among the key topics to which coders were sensitized were statements of praise or criticism of the factions fighting for power. The results led to tentative conclusions about the level of support for different factions in the different regions of China.

The single most significant weakness of content analysis is locating messages relevant to the research question. Some topics do not appear with any regularity in the available media. Even more difficult to find in the same medium are combinations of two or more variables in ways that permit tests of their relationship. Content analysis tests the imagination and creativity of the researcher as much as any research technique.

Another important limitation of content analysis is that it cannot be used to test causal relationships between variables. Researchers and consumers of research must resist the temptation to infer such a causal relationship. For instance, it has been shown that the portrayal of crime on American television in the 1960-1980 period has increased at about the same rate as did actual crime in the United States. However, the correspondence between the two rates does not necessarily mean that violence on TV has caused crime to increase in American society. It may or may not have, but content analysis cannot support or refute such a hypothesis. Content analysis must be combined with other research strategies to demonstrate cause and effect.

An example of the use of a multimethod approach to determine cause and effect is the research effort to demonstrate effects of sexually explicit material in movies, television, and the printed media on the sexual behavior of those who saw the films. The Pornography Commission reviewed content analyses of movies, TV, and magazines to determine the prevalence of sexually explicit material. The next step in the scientific assessment of pornography was to survey the public to determine its exposure to the

sexually explicit material and to conduct experiments about the effects of such material on people's feelings and behavior. Surveys of various populations provided estimates of the size of the audience for pornographic television programs and movies. Laboratory experiments exposed college students to sexually explicit materials and determined the degree of sexual arousal attributable to such exposure. The content analysis, survey research, and experimental research, in combination, provided a much more comprehensive answer to the question of the impact of pornography in American society than was possible from the use of any one of these research strategies. The content analysis revealed the nature of the sexual material available, the surveys provided estimates of the incidence of exposure to it, and the experiments illustrated some of the effects of the material on sexual arousal.

A potential problem for content analysts is that the material recorded and saved, and therefore available for analysis, is not representative of all such material. Newspaper editors apply their paper's political philosophy in determining whether an article is published and in editing those that do appear. Many hours of videotaping are needed to produce a thirty-minute evening news program, and the news editor selects which stories appear and what material is used in their presentation. Videotapes of events not reported as well as unused portions of tape are erased, and thus only a nonrepresentative segment of the material collected is saved. The bias of the sender and the nonrepresentative nature of the selection and retention process are a significant and often insoluble problem. The existing data, biased as they are, are all the researcher has. He or she must try to identify biases in the messages available and, if possible, locate other sources. However, often all that is possible is the frank admission that "this is all there is" and that biases, both those noted and those unknown, are unavoidably present.

The advantages of content analysis must be weighed against its disadvantages and against alternative research strategies. It is appropriate for some problems but not others. Content analysis is useful in studying trends over time, particularly in trying to reconstruct events in the past. It is helpful in assessing events and processes in social groups where the researcher is denied personal access but may accumulate records or communications. Content analysis is also a useful strategy in exploratory research. It definitely is not a substitute for experimental research in testing for cause-and-effect relationships.

III. SOURCES OF MATERIAL

Communications available for content analysis are almost unlimited. Personal documents such as letters and diaries are an excellent source for some kinds of information. W. I. Thomas and Florian Znaniecki (1971) studied

the adjustment of Polish immigrants in America at the turn of the century through the content analysis of 750 letters sent from immigrants in Chicago to their relatives.

The printed media—newspapers, magazines, books, and pamphlets—are good sources of communications and are readily available in libraries. Back issues of most major newspapers and magazines are available so that trend analysis is possible. An example of such work is Gecas' (1972) study of aggressive behavior in popular fiction as revealed in an analysis of short stories in four magazines—*Esquire* (middle- and upper-class men's magazine), *McCall's* (middle-class women's magazine), *Argosy* (lower-class men's magazine), and *True Confessions* (lower-class women's magazine)—between 1925 and 1965. Gecas randomly selected two issues of each magazine per year, took one short story from each magazine, and coded the content into seven categories of aggressive behavior: physical aggression, verbal aggression, economic sanctions, social sanctions, legal sanctions, hurting a person by injuring a loved one, and fantasy aggression. Gecas' analysis revealed that aggression in popular fiction had changed little over the 40-year period.

Textbooks have frequently been analyzed to assess levels of prejudice (negative stereotypes) and discrimination (neglect of minorities' contribution to history) as evident in materials used in the public schools (Bowker, 1972).

Systematic analysis of advertisements in newspapers and magazines also yields insight into national culture. Condie and Christiansen (1977), for example, examined advertisements for hair strengtheners and bleaching creams in *Ebony*, a popular black magazine, between 1949 and 1972. They discovered a decline (fewer advertisements) in the use of both hair straighteners and bleaching creams in the 1960s, which they interpreted as evidence for a growing "Black is Beautiful" movement. Black Americans during this period, at least those represented in magazines for black people, seemed to emphasize Afro hair styles and dark skin color as positive traits.

Cartoons have been analyzed for information about social trends or movements. Houts and Bahr (1972) analyzed cartoons in the *Saturday Evening Post*, a popular magazine, from the year 1922 through 1968, when the *Post* ceased publication as a weekly. They were interested in the stereotypes of blacks and Indians presented in cartoons over this 46-year period. They discovered very little change in the cartoonists' representation of Indian Americans, but black Americans almost disappeared from cartoons after 1960. They were replaced by black Africans, typically drawn with bones through their noses, and a white hunter in the village cooking pot.

Television, radio, and movies have been a fruitful source of material for content analysis. Numerous studies have explored topics such as women's roles, stereotypes of minorities, sexual behavior, and violence in television and movies. The titles and lyrics of popular songs have been analyzed

for insights into the values of young people. Denisoff (1975) reviewed studies of rock music as a window to the world of pop culture.

Sermons delivered in Middletown, U.S.A., during 1924 and 1978 were analyzed to assess whether emphasis on secular topics, including social and political problems, had increased during the past 50 years (Lynd and Lynd, 1929; Caplow, Bahr, and Chadwick, 1983). Titles and summaries were obtained for 109 sermons given by Middletown ministers in 1978, and they were coded into one of two categories: religious or secular. The results revealed that the sermons of 1978 were as religiously oriented as were those of the early 1920s.

A researcher may ask people to write a life history or an essay describing their values, which can then be analyzed. Such a technique is obtrusive in that respondents are asked to write a communication especially for the researcher, but it is an excellent source of data for content analysis.

IV. COMPUTER ASSISTED CONTENT ANALYSIS

As in much data analysis, computers are a valuable tool in content analysis. When the unit of analysis is a word or set of words, the coding and counting can be done by computer. Computer programs like the "General Inquirer" identify, within text, those words and phrases that belong in specified categories. In addition, such programs count the frequency of identified words, graph the output, and compute test of significance (Stone et al., 1966). Specialized computer dictionaries have been developed for most of the social sciences which relate a large number of words to particular thematic categories. Combinations of words can also be specified as belonging to a given category. Thus thousands of words or phrases can be coded into twenty or fewer categories (Dunphy, 1966, Carney, 1972). Computers will play an increasingly important role in content analysis as optical scanners that can read text come into general use, as well as voice scanners that can "listen" to text and thus provide more convenient input.

Although computers facilitate content analysis when the unit of analysis is a word or phrase, they are limited in thematic or topical searches. When text must be searched for ideas or thoughts, computers are not as sensitive as human coders. Also, if the dross rate is high, as when much text must be reviewed to produce only a few useful "units" of data, having the entire text to input may be more costly than hiring people to scan it. Improved optical and video scanners may eventually change the cost-effectiveness of computer scanning as compared with visual scanning by researchers. However, computer technology has a long way to go before computers will replace trained researchers in tasks such as watching television programs to analyze their content.

V. DOING CONTENT ANALYSIS

Statement of the Problem

Content analysis begins with a specific proposal, a systematic statement of the problem to be studied. As an example, let us take the work of Wales and Brewer (1976), who wanted to know whether differences in sexual expression noted by Kinsey in the late 1940s and early 1950s were inherent male-female differences or if the differences reflected a greater social control of female sexuality.

Selection of Communications

The researcher must locate a source of communications relevant to the research question. A thorough examination of one's home, office, and public library will frequently yield usable communications. The time period to be studied must be determined, and if the number of relevant communications during that period is excessive, a sample can be drawn.

In the search of gender differences in sexual expression, Wales and Brewer (1976) elected to study graffiti in high school restrooms. They selected four high schools in a midwestern city: a lower-class black school, a lower-middle class white school, a middle-class integrated school, and an upper-class white school. The researchers copied on 3 x 5 cards all the graffiti in both men's and women's restrooms in these four high schools.

Operational Definitions

The first step in developing operational definitions is to select the unit of analysis. The unit can range from a single word to an entire message. If it is a single word, the number of times that word appears can be counted. Thus the frequency with which words like "religion," "God," "heaven," and "hell" appear in a student newspaper might be used as an indicator of student religiosity. At the other extreme, the single most important theme in an entire book, movie, TV program, or article can be the unit of analysis. Between these two extremes, sentences, paragraphs, scenes, and ideas may serve as the unit.

The selection of a **unit of analysis** is made with reference to the topic and type of communications studied. Preliminary examinations of the communications in small-scale pilot studies will suggest possible categories into which material can be coded. Usually one experiments with several category systems before developing a set that coders can use to provide the classifications needed for the study. Categories must be described in enough detail that different coders will categorize the data into the same number of units, distributed in almost the same way. Categories must be mutually exclusive so that a word, paragraph, or theme belongs in one and only

one category. Also, the categories must be exhaustive, so that all units examined fit in an appropriate category. Sometimes a miscellaneous or residual category is added for units that occur rarely or are uncodable for other reasons.

In the study of graffiti, the unit of analysis was a "thought." Thus a single word or a five-line rhyme was counted as a unit. Chained responses, that is, graffiti written in response to other graffiti, were treated as a single unit. Wales and Brewer developed an initial set of categories on the basis of previous studies but were forced to modify their category system when they coded the graffiti they had collected. They finally developed the following sixteen categories (Wales and Brewer, 1976: 118-119).

1. Racial Insults—insults excluding the use of sexual or scatological words
2. Sexual Insults—any insult using sexual or scatological words excluding reference to race
3. Racial/Sexual Insults—combined use of sex and race
4. General Insults—any insult without use of racial, sexual, or scatological words
5. Sexual Humor—any attempt at humor using sexual or scatological words
6. General Humor—humor excluding sexual or racial references
7. Sexual Request—names, phone numbers
8. Sexual or scatological words—any single word, broken phrase
9. Romantic—any statement of attachment excluding sexual
10. General Racial—any reference to race excluding sexual words
11. Political—any political reference excluding race and sexual words
12. Drugs—any reference to drugs or drug usage
13. Religion—any religious reference
14. Morals—any moralistic statement
15. Names—no content other than use of names or initials
16. Miscellaneous—no categorizable content

Training Coders and Checking Reliability

Those who code the content of messages into categories must be trained to recognize the salient features which determine the appropriate category for a unit. Ideally two or more coders should independently code the same set of messages and their reliability should be checked by an item-by-item comparison of their work. The simplest way to compute reliability between coders is to calculate a **Coefficient of Reliability** by dividing the number of units placed in the same category by the number of units coded.

$$\text{Coefficient of Reliability} = \frac{\text{Number of units in same category}}{\text{Total number of units coded}}$$

Holsti (1969) reviewed several other ways to compute reliability which take into account the number of categories and the probability that each category would be used. Although such sophisticated techniques provide a more accurate measure of reliability, the easily calculated Coefficient of Reliability is widely used and generally adequate. Although there is no absolute standard of reliability demanded, a widely accepted threshold is 60 percent. If there is less than 60 percent agreement between coders, then the operational definitions probably need to be made more specific. Sometimes the problem is that two or three categories are so similar that the coders can't differentiate between them. This problem can be solved by combining such overlapping categories into one.

The graffiti study used three judges or coders who independently coded the graffiti. The operational definitions worked very well, as the reliability among coders was .96.

Analysis of Data and Writing of Report

The quantitative data obtained by content analysis can be analyzed with standard statistical techniques. If desired, tests of significance and measures of association can be calculated between different groups, time periods, and so on.

The graffiti data were analyzed to determine if there were significant differences in the amount and type of graffiti written by males versus females, blacks versus whites, and lower-class versus middle-class students. The most striking finding was that females accounted for 88 percent of all written material. In all four high schools, girls outproduced the boys in volume of graffiti by a ratio of 3.5 to 1. As expected, the females wrote most of the romantic inscriptions. A finding that had not been expected was that the girls from upper-class backgrounds specialized in erotic messages. These findings refuted Kinsey's notion of inherent differences between males and females in sexual expression; the girls in this study were more than a match for the boys, at least in such written expressions as are found in public restrooms. The results were published in *The Journal of Social Psychology* (Wales and Brewer, 1976).

VI. EXEMPLARY CONTENT ANALYSIS STUDIES

Violence Profile

The original purpose of the Violence Profile study was to provide periodically an estimate of the level of violence contained in television programs (Gerbner et al., 1980). This ongoing project, begun in 1967, has since branched out into other areas of programming, but we will limit our discussion here to violence. The latest violence profile available is for 1980—

1981 (Gerbner, 1981). These profiles are used by the networks, government officials, and other interested parties to monitor TV content.

Selection of communications Gerbner and his associates selected the three major networks, ABC, CBS, and NBC, as their source of communications. These three networks probably accounted for over 95 percent of the television audience in 1967. New networks, especially the cable networks, now have a larger share of the audience and it is speculated that eventually Gerbner and his associates will include them in their analyses. The hours of programming studied were prime time, 8 to 11 P.M. Eastern Standard Time, and children's viewing time, 7:30 A.M. to 2 P.M. on Saturday and Sunday. A sample of the viewing hours was selected by collecting data for one week each year. It was assumed that a single week included all of the series and a representation of specials, movies, and other programs.

Operational definitions The unit of analysis was a violent act by a character. A violent act was operationally defined as "an overt expression of physical force (with or without a weapon against self or others) compelling action against one's will on pain of being hurt and/or killed or threatened to be so victimized as part of the plot" (Gerbner et al., 1980: 11). Accidental aggression was included but verbal abuse and threats were not.

The units observed were combined in a Violence Index (VI). The component of the VI included: percent of the programs with at least one violent act (%P); the rate of violence acts per program (RP) and per hour (RH); the percent of characters who were either perpetrators or victims of violence (%V); and the percent who were killed or who participated in killing (%K). These five measures, added together, produced the Violence Index. Because of the perceived importance of the rates of violence, both rates were doubled in calculating VI. The final formula was as follows:

$$\text{Violence Index} = \%P + 2RP + 2RH + \%V + \%K$$

In addition, several other characteristics, including the type of program, and the sex, race, occupation and marital status of major and minor characters, were coded so that violence by type of program and participant could be identified. These characteristics were also used in analysis about programming in addition to violence. For example, the marital status of the characters was coded into the following categories (Signorielli, 1982: 588):

1. Single—included characters who cohabited with the opposite sex but did not plan to marry
2. Married—included those who were presently married, got married, or planned to be married

3. Formerly Married—including widowed, divorced, and separated
4. Mixed—including characters who fell in more than one of those categories
5. Cannot Code—the story did not provide sufficient information about marital status

The last category is an example of a miscellaneous or residual category for units of analysis that do not fit into other specific categories.

Training coders and checking reliability Gerbner recruited 12 to 18 coders each year who were trained for three weeks. They worked in pairs viewing the tape-recorded programs. Two pairs were assigned to independently code each program and their results compared for reliability. Reliability varied for the different characteristics. It was very high for sex (about 98 percent) and significantly lower for more subjective characteristics like motivation and personality traits (about 50 percent). Over the years the reliability for the violence coding has been 75 to 85 percent, and no results have been presented where coder reliability was 60 percent or lower (Gerbner et al., 1980).

Analysis of data and writing of report Violence Indexes were calculated for both prime-time and children-time viewing. During prime time in the 1980–1981 season, 73 percent of the programs analyzed had at least one violent act, and the average rate of such actions was six per hour. Also, over half (52 percent) of all major characters were involved in violence, either as aggressor or victim.

The violence in programming during the prime viewing hours for children was even higher. Almost every program (97 percent) contained at least one violent episode, and the rate of aggressive acts was 25 per hour. In addition, 88 percent of all major characters were involved in violence.

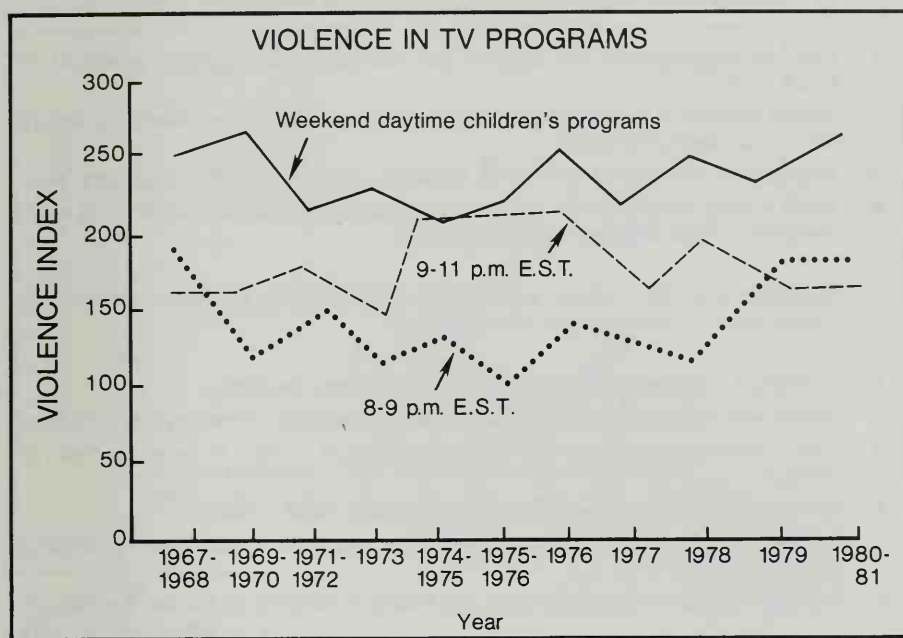
The trends in Violence Index scores for the years 1967–1979 are presented in Illustration 9.2. Although there have been minor fluctuations in the violence index over the years, the general trend has been a steady, high rate of such material, especially in programming for children.

Gerbner and his associates have published numerous articles reporting their findings about violence in television programming. For additional detail and references to other reports, see Gerbner and Gross, 1976, and Gerbner et al., 1980.

The Emergence of Indian Activism

During the 1960s and early 1970s several protest activities by Indian Americans captured national attention. The fish-ins in Washington State, the seizure of Alcatraz, the Trail of Broken Treaties and the associated

ILLUSTRATION 9.2 Violence Index 1967-1980.



occupation of the Bureau of Indians Affairs' offices, and the occupation of Wounded Knee, South Dakota, are probably the most well known. Day (1972) wanted to study the emergence of collective actions fostering tribalism as well as disruptive protest activities by Indian Americans. He analyzed newspaper reports to estimate the number of collective activities by Indian Americans between 1961 and 1970.

Selection of communications Day (1972) selected the *New York Times* as the medium for study because using a single national newspaper simplified the task of collecting and coding the communications. He acknowledged the potential bias of using a single newspaper but decided that the *New York Times'* national coverage was adequate for his purposes. He reviewed all issues of the *Times* published between January 1961 and September 1970.

Operational definitions The unit of analysis was a report of collective action by Indian Americans. The actions were coded into two major categories, obstructive and facilitative. There were four subcategories of obstructive and six of facilitative actions. The operational definitions he used were as follows:

Obstructive activities: those activities that blocked or halted ongoing activities by violating laws, regulations, or strongly held norms and values

1. Delayed ongoing activities such as dam construction, logging operations, or beach use
2. Seized control or occupied government offices, military installations, fishing sites, and similar facilities
3. Engaged in nonviolent picketing, speeches, sit-ins, marches, and boycotts
4. Made public verbal attacks against government officials such as jeering at the Secretary of the Interior during a speech

Facilitative activities: those actions that used existing political and social institutions to promote tribal development

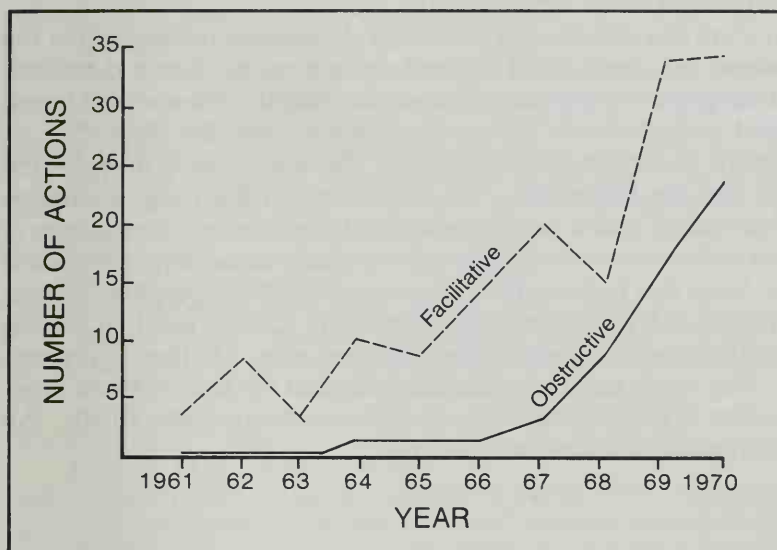
1. Promoted economic development of reservation resources
2. Sponsored public relations projects to reduce negative stereotyping of Indians
3. Held conferences to organize different tribes or groups to increase their influence
4. Arranged new educational institutions under Indian control
5. Initiated legal proceedings to reassert land, mineral, water, grazing, fishing, and civil rights
6. Initiated studies or inquiries into the living conditions of Indian Americans

Specific incidents were counted only once, no matter how many times they were reported in the *Times*. The only exception was that the occupation of Alcatraz lasted for several months, and distinct major events there were counted separately.

Training coders and checking reliability Day did all the coding himself and thus did not train anyone, nor did he check the reliability of his own work. His findings would be more creditable had there been an independent coding of the newspapers, or at least a sample of them, to establish a reliability rate.

Analysis of data and writing of report Day's findings, presented in Illustration 9.3, revealed that both facilitative and obstructive types of activities had increased during the decade, with obstructive actions less frequent earlier in the decade but rising sharply after 1966. The increase in collective action is not surprising, given the poverty and deprivation experienced by Indian Americans both on reservations and in the cities. What is surprising is that there have been so few obstructive tactics by Indian Americans. Nothing in Indian country has approached the degree of violence that characterized the urban riots of other minority groups during this same decade.

ILLUSTRATION 9.3 Collective actions by Indians 1961–1970.



VII. SUMMARY

Content analysis is the study of communications to describe social behavior or to test hypotheses about relationships revealed in such communication. The content of messages is analyzed to determine how often selected words, ideas, or themes appear and how large a portion of the total communication they make up. The results permit the analyst to make inferences about characteristics of the sender or the possible impact of the messages on the audience. Television and radio programming, movies, books, newspapers, letters, magazines, songs, responses to open-ended questions, and projective psychological tests are potential data for content analysis.

One advantage of content analysis is that it is nonreactive; it does not impose upon respondents or subjects. Generally the recorded communications are readily available to the researcher. In addition, compared with other types of research, it is relatively inexpensive. Another advantage of content analysis is that it can be used to study populations to whom the researcher cannot gain direct access. Newspapers, radio and television broadcasts, and similar public communications can be analyzed for clues about what is happening within a closed community.

The most difficult problem with content analysis is finding messages relevant to the topic the researcher wants to study. A second significant limitation is that content analysis cannot establish causal relationships. In addition, there are potential biases in the communications themselves;

there are selection biases in that certain communications have survived in archives or libraries while others are lost.

Computers can rapidly and accurately do content analysis when the unit of analysis is a word or set of words which can be clearly identified. It is difficult at present to program computers to identify themes and ideas.

Content analysis starts with a statement of what the researcher intends to study. Communications are then identified which describe the behavior or test the hypothesis. The time frame of the communications must be determined and a sample selected if the number of messages is large. The selection of a unit of analysis is necessary to develop operational definitions. Once this is done, then clear instructions are prepared so that coders or judges can reliably recognize the units. Coders must be trained to extract the information from the messages and place it in the appropriate categories. The communications should be coded by two or more independent coders to demonstrate the reliability of the analysis. Finally, the data are analyzed and the report prepared.

REFERENCES

- ARONOW, E., and M. REZNIKOFF, 1976. *Rorschach Content Interpretation*. New York: Grune & Stratton.
- ATTKISSON, C. CLIFFORD, LEONARD HANDLER, and RAYMOND R. SHRADER, 1969. "The Use of Figure Drawing to Assess Religious Values." *The Journal of Psychology* 71: 27-31.
- BOWKER, LEE H., 1972. "Red and Black in Contemporary American History Texts: A Content Analysis." *Native American Today*, ed. Howard M. Bahr, Bruce A. Chadwick, and Robert C. Day. New York: Harper & Row.
- CAPLOW, THEODORE, HOWARD M. BAHR, and BRUCE A. CHADWICK, 1983. *All Faithful People*. Minneapolis: University of Minnesota Press.
- CARNEY, THOMAS F., 1972. *Content Analysis*. Winnipeg, Canada: University of Manitoba Press.
- CHADWICK, BRUCE A. and ROBERT C. DAY, 1972. "Response Hierarchy: Influence Tactics in a Social Exchange Relation." *Journal of Social Psychology* 86 (March): 241-246.
- CHAI, TRONG R., 1977-78. "A Content Analysis of the Obituary Notices on Mao Tse-Tung." *Public Opinion Quarterly* 41: 475-487.
- COMMISSION ON OBSCENITY AND PORNOGRAPHY, 1970. *The Report of the Commission on Obscenity and Pornography*. New York: Bantam Books.
- COMSTOCK, GEORGE A., and ELI A. RUBINSTEIN, 1972. *Television and Social Behavior*, Vol. I-V. Technical Reports to the Surgeon General's Scientific Advisory Committee on Television and Social Behavior. Washington, D.C.: U.S. Government Printing Office.
- CONDIE, SPENCER J., and JAMES W. CHRISTIANSEN, 1977. "An Indirect Technique for Measurement of Changes in Black Identity." *Phylon* 38, No.1: 46-54.
- DAY, ROBERT C., 1972. "The Emergence of Activism as a Social Movement" in

- Native Americans Today*, ed. Howard M. Bahr, Bruce A. Chadwick, and Robert C. Day. New York: Harper & Row.
- DENISOFF, R. SERGE, 1975. "Content Analysis: The Achilles Heel of Popular Culture?" *Journal of Popular Culture* 9: 456-460.
- DUNPHY, P. C., 1966. "The Construction of Categories for Content Analysis Dictionaries" in P. J. Stone, et al., *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge, Mass.: The MIT Press.
- GECAS, VICTOR, 1972. "Motives and Aggressive Acts in Popular Fiction: Sex and Class Differences." *American Journal of Sociology* 77, No. 4: 680-696.
- GERBNER, GEORGE, 1981. Testimony Before the Subcommittee on Telecommunications Consumer Protection and Finance of the Committee on Energy and Commerce, U.S. House of Representatives, Washington, D.C., October 21. Mimeocopy.
- GERBNER, GEORGE, and LARRY GROSS, 1976. "Living with Television: The Violence Profile." *Journal of Communication* 26 (Spring): 173-199.
- GERBNER, GEORGE, LARRY GROSS, MICHAEL MORGAN, and NANCY SIGNORIELLI, 1980. "The 'Mainstreaming' of America: Violence Profile No. 11." *Journal of Communication* 30, No. 3 (Summer): 10-29.
- HOLSTI, OLE R., 1969. *Content Analysis for the Social Sciences and Humanities*. Reading, Mass.: Addison-Wesley.
- HOUTS, KATHLEEN C., and ROSEMARY S. BAHR, 1972. "Stereotyping of Indians and Blacks in Magazine Cartoons" in *Native Americans Today*. Howard M. Bahr, Bruce A. Chadwick, and Robert C. Day (eds.) New York: Harper & Row.
- HOWES, RICHARD J., 1981. "The Rorschach: Does It Have a Future?" *Journal of Personality Assessment* 45, No. 4: 339-351.
- LYND, ROBERT S., and HELEN MERRELL LYND, 1929. *Middletown*. New York: Harcourt, Brace & World.
- SANDERS, WILLIAMS (ed.), 1974. *The Sociologist as Detective*. New York: Praeger.
- SIGNORIELLI, NANCY, 1982. "Martial Status in Television Drama: A Case of Reduced Options." *Journal of Broadcasting* 26, No. 2 (Spring): 585-597.
- SILVERMAN, L. THERESA, JOYCE N. SPRAFKIN, and ELI A. RUBINSTEIN, 1979. "Physical Contact and Sexual Behavior on Prime-Time TV." *Journal of Communication* (Winter): 33-34.
- STONE, P. J., D. C. DUNPHY, M. S. SMITH, and D. M. OGILVIE, 1966. *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge, Mass.: The MIT Press.
- THOMAS, W. I., and FLORIAN ZNANIECKI, 1971. *The Polish Peasant in Europe and America*. New York: Octagon Books (first publication in 1919).
- WALES, ELIZABETH, and BARBARA BREWER, 1976. "Graffiti in the 1970's." *The Journal of Social Psychology* 99: 115-123.

Secondary Analysis

ten

I. Introduction

II. Advantages and Disadvantages of Secondary Analysis

Advantages

Disadvantages

III. Doing Secondary Analysis

IV. Illustrative Examples

Secondary analysis of experimental data: the Hawthorne data

Secondary analysis of survey data: correlates of widowhood studies

Secondary analysis of evaluation research: the Bureau of Indian Affairs relocation program

V. Summary

I. INTRODUCTION

Secondary analysis usually refers to the use of research materials by persons other than those who gathered them and/or for purposes different from the original project objectives. The most successful example of secondary analysis is demography, which has become a scientific discipline in its own right. Demography takes official statistics—frequently census materials—collected by government agencies for administrative purposes and uses them to draw generalizations about fertility, mortality, migration, occupational change, and other processes far beyond those intended by the agencies collecting these materials.

Because we are talking of the social sciences generally, including history, it is necessary to distinguish secondary analysis from the conventional classification of documents and other types of data into primary and secondary source materials. In the conventional classification, **primary sources** include artifacts or statements of participants and eyewitness observers, or accounts by nonobservers based on reports or field notes of participants who are no longer available. **Secondary sources**, clearly less reliable, include accounts by nonobservers based on existing information not directly examined, or reports based on accounts by nonobservers who did not publish primary data. (Naroll, 1962: 31–32) Some writers make a simpler and sharper distinction between primary and secondary sources, distin-

guishing between eyewitness reports (primary) and other compilations of documentary evidence (secondary sources) (Bailey, 1978: 290–291).

In the conventional classification of source materials, primary materials are clearly more desirable. *Secondary analysis*, however, is intrinsically neither better nor worse than *primary analysis*. A secondary analyst may produce a more accurate report than did the researcher who initially collected the data. In fact, except for situations in which the documentation of a data set is insufficient to allow secondary analysts necessary insights into problems of data collection, there is every reason for the secondary analysis to be as complete and accurate as that done by the original researchers. The secondary analyst, after all, has the same codebooks and data files to work with. The secondary analyst may even have a slight advantage because he or she has not devoted substantial personal and project resources to the costly stages of study design and data collection. Presumably he or she can then devote more time to analysis, interpretation, and creative thought about what the data show and what they mean.

Some writers define secondary analysis more strictly, and others more loosely, than we have. Moser and Kalton (1972: 43) simply refer to "re-analysis from a different standpoint of the findings in someone else's research." Simon (1978: 379) talks of "data dredging" as the examination of data collected for a particular purpose for possible new relationships, warns that such relationships may not be tested with the same statistical procedures that apply to standard primary research, and concludes that when one has dredged up a pattern via secondary analysis, one must test its existence in different data (Simon, 1978: 450–454).

Kendall and Lazarsfeld (1950: 134) distinguished between primary analysis, which was determined by the original purpose of the study, and secondary analysis, in which findings were derived from surveys designed for other purposes. They added that secondary analysis had some standard rules or procedures but that they were rarely articulated. This view is particularly interesting in that, from today's standpoint, secondary analysis was little used and its methods largely undeveloped and unappreciated in 1950. In the same paragraph, Kendall and Lazarsfeld described the first two volumes of *The American Soldier* (Stouffer, Suchman et al., 1949; Stouffer, Lumsdaine et al., 1949) as "a text which not only demonstrates the merits of secondary analysis, but which is also so rich that some of the rules of analysis can be extracted and exemplified."

Although their article and others in the same volume (Merton and Lazarsfeld, 1950) did make explicit some rules for secondary analysis, it was 22 years before a manual designed to teach students the elements of a scientific secondary analysis of survey data was published by Herbert Hyman (1972). In it Hyman dates the beginnings of interest in greater utilization of data from past surveys at about 1957, when a proposal to establish a library for survey data was submitted to the Ford Foundation

and when the Roper Public Opinion Research Center at Williamstown, Massachusetts, was officially organized as a general archive (Hyman, 1972: 1). However, Hyman himself was teaching a course at Columbia University entitled "Methods and Applications of Secondary Analysis" in 1951. In our view, the identification of secondary analysis as a viable method dates from the work during the late 1940s of Stouffer and his associates and the subsequent critique and reanalysis (Merton and Lazarsfeld, 1950) of some of the findings from the first two volumes of *The American Soldier*. Hyman's (1972) book remains the best single manual on how to do secondary analysis. However, some recent more specialized works are also very useful. These include a collection of papers on the techniques of the secondary analysis of **evaluative research** (Boruch, Wortman et al., 1981) and a special issue of the *Journal of Social Issues* (No. 1, 1982) devoted to various techniques vital to the study of women and social change.

II. ADVANTAGES AND DISADVANTAGES OF SECONDARY ANALYSIS

Like other research methods, secondary analysis has assets and liabilities. In our view, its benefits will usually far outweigh its liabilities for research questions amenable to secondary analysis. Ultimately, of course, that decision belongs to the individual researcher, those who sponsor research, and the consumers of research.

Advantages

The advantages of secondary analysis from the standpoint of both the sponsor of research and the individual independent scientist were spelled out more than two decades ago by Barney Glaser (1962, 1963). An important benefit noted by Glaser (1962: 73-74) was the economy in money and time. It is cheaper to reanalyze existing data than to collect and analyze new data, and findings are much more rapidly generated by such reanalysis than through the drawn-out process of designing a new project and collecting new data. Glaser also argued that sensitivity to the cost efficiencies of secondary data might predispose potential clients favorably toward social research. The major costs of research are those associated with data collection, and the secondary analyst may totally avoid these. Analysis of data is relatively inexpensive, and access to computers and modest secretarial assistance may often be obtained even by the lone scholar. There is also a societal benefit in the sense that the primary investigators use only a part of the data collected at such great expense, and storing or discarding the original data wastes societal, organizational, and intellectual resources.

Significant social benefits of secondary analysis include the fact that it is much less obtrusive than primary research, and there may be times

when trying to study a community by standard primary methods would heighten existing tensions. A related benefit is that the good will of potential respondents is not unlimited, and in the interest of efficiency and respect for the time and privacy of one's fellows, intrusions for primary data collection should be minimized. Secondary analysis increases the likelihood that a previous intrusion will have been worthwhile, that the time respondents devoted to a previous survey—in no sense a free resource—will yield some positive results.

Glaser (1963) urged more attention to secondary analysis by the "independent researchers" whose motivation for research was to satisfy intellectual curiosity and who wished at the same time to contribute to the accumulation of scientific knowledge. Independent researchers frequently work alone, have limited financial resources, and typically are part-time researchers whose professional assignments include other duties.

Indeed, there may be a natural division of labor between the data collector and the analyst. Especially in large-scale studies, Glaser (1963: 11) argued, these tasks were often best accomplished by teams of specialists. Twenty years later Norval Glenn made the same point in a review of the General Social Surveys conducted by the National Opinion Research Center and made available to anyone who wants to analyze them and can pay the very modest costs of copying codebooks and purchasing data tapes. Glenn said that with large quantities of national or regional data available to all interested researchers,

... an almost revolutionary change in survey research would seem to be occurring. Until recently, survey data were analyzed primarily by the persons who designed the surveys, but there seems to be a rather strong trend toward separation of survey design from data analysis. One can almost envision a time when some survey researchers will specialize in survey design and others will specialize in data analysis. (Glenn, 1978: 533)

The advantage of secondary analysis by the independent researcher is that a fresh, creative approach and personal style is likely to be applied to the data. The personality of the independent researcher is often different from that of "bureaucratic" researchers and the idiosyncracies of such people are sometimes the qualities needed to provide a perspective different from that of the more conservative, possibly less innovative group researchers. The very personality characteristics that might make the independent researcher a poor team member may make his or her insights and unusual viewpoints particularly valuable to the scientific enterprise. The independent researcher subjects the data and interpretations to the critical judgment of an "outsider" who has nothing to lose, and may with impunity disagree with the "majority view" about what the data mean (Glaser, 1963: 12).

Glaser identified three types of independent researchers. First is the

teacher, who contributes an "economy of interest" by bringing a deep knowledge of a specialty, and of the research methods most appropriate to that specialty, to bear on a data set, thereby enhancing the possible yield of knowledge.

A second type of independent researcher is the *student*, who usually lacks resources and whose apprenticeship in research and writing may produce little that advances the scientific enterprise or his or her own career. This outcome is especially probable if a student's research is limited to data collected by the student. Glaser recommends that in dissertation and thesis projects, students be encouraged to work with secondary data sets of high quality so that the results of their apprenticeship not only may be means to achieving an academic degree but may be scientific contributions of value.

Finally, Glaser identifies the *otherwise employed*, the person in an occupational setting where research is not a part of the job description. Some of the "otherwise employed" have an interest in doing research for personal satisfaction, for professional advancement, or to maintain a sense of identification with the scientific discipline in which they were trained. Secondary analysis provides a vehicle whereby such people, at minimal financial or other costs, can keep up with or make a contribution to a field. If their research involves the analysis of large, high-quality data sets, their talents may contribute advances to scientific knowledge as a whole that would be impossible if they were limited to data they could collect by themselves.

Hyman (1972: 10–24) cites benefits for theory and substantive knowledge from secondary analysis as a strength of this type of research. The secondary researcher searches through a wide array of materials spread over geographic space and time which may yield greater scope and depth than is usually possible in the single primary research project. Such benefits include, first, an improved understanding of history via the analysis of data collected in the past, and second, a heightened opportunity to understand change by the stringing together of studies done at different times which dealt with comparable issues. Note that this technique amounts to **longitudinal research** in reverse, for instead of making a baseline measurement and then continuing to monitor change for 20 or 30 years into the future, the researcher can discover baselines in studies conducted 20 or 30 years in the past and, with effort and luck, locate other studies done in the intervening period with comparable items, thereby producing "instant" longitudinal data for a decade or two. A primary research project may be combined with the data from the earlier studies to provide contemporary follow-up.

A second advantage for theory and substantive knowledge that secondary analysis provides is the opportunity to examine problems on a comparative basis—not only over time but in different societies at approximately the same time. Cross-national studies are very difficult to accom-

plish, but many have been done, and in secondary analysis, researchers can try to make the most of the data accumulated on comparable variables from two or more ethnic or national settings.

A third benefit to theoretical and substantive knowledge that secondary analysis may provide is the extension of reliable knowledge through *enlargement*, or the combination of several studies to obtain a very large number of cases. The opportunities to study rare or deviant characteristics is greatly enhanced by drawing such cases from many different studies. For example, a primary research project large enough to produce a sizable sample of interracial marriages might be beyond the scope of most researchers. However, at least with respect to simple demographic characteristics, it may be possible to collect data on the handful of interracial marriages contained in several dozen political polls, or other prior studies, and eventually to accumulate a large enough sample of persons whose marriages are interracial to permit some reliable generalizations.

A fourth benefit Hyman (1972: 21–23) discusses is simple **replication**. That is, the use of secondary analysis may allow a scientist to find a few or perhaps many studies that have included a set of variables the investigator wishes to study. A pattern or model observed in one's own primary data or in a secondary analysis study gains increased reliability as it appears in successive studies. Rather than do the successive studies personally, the secondary analyst draws upon the research of others and, in effect, does "retrospective replication" that may produce results just as reliable as findings from a series of primary studies that were designed for purposes of replication.

Hyman notes the advantage of elevating and enlarging theory that often accompanies secondary analysis. That is, because several studies may not use precisely the same **indicators** for a given concept, the analyst is forced to reexamine the concept and to ask whether indicators worded differently may, in fact, be tapping the same underlying concept. The researcher often is forced to move to a higher level of abstraction to manage the diversity of indicators and variations of **operationalization** that appear in several data sets. This sort of abstract thought is as likely to lead to scientific breakthroughs as is the deductive thought that leads a researcher from the realm of general theory to the narrow hypothesis. The secondary analyst, because he or she examines many studies conducted in different ways, at different times, and in different places, acquires a skill in dealing with ambiguity that permits him or her to find common relationships in studies that at first seem to be markedly different. This ability to view data sets from widely differing contexts, to look for unifying principles or procedures that permit the analyst to make sense out of what seem to be disparate, unrelated bits and pieces, represents the sociological imagination in its finest form.

Disadvantages

The most frequent problem encountered by the secondary analyst is locating data relevant to a particular research question. In spite of the many thousands of studies in data archives, it is sometimes difficult to find ones with the variables the researcher is interested in. Sometimes the relevant data are not accessible because the original researcher has not made them available. As was mentioned earlier, creativity in locating data sources and in reconceptualizing variables is critical to secondary analysis.

Although thousands of studies are in data archives, the original data for most published social science research are not available to researchers because the investigators have not released the data for public use, have destroyed them, or have not prepared and retained them in forms suitable for archiving and secondary analysis.

Other reasons why research data are not more readily available include an alleged (Bryant and Wortman, 1978: 382) belief among many social scientists that their research is irrelevant to pressing social issues and hence not essential to retain in archives for reanalysis. A related belief contributing to complacency toward secondary analysis among some social scientists is that the proper test of the validity of research findings is their replication, not reanalysis of the data set that produced the findings in the first place.

There is also resistance among original investigators to making their data available for "pragmatic or personal reasons" that include concern that reanalysis of their data may reveal methodological or statistical errors in the reports, that proprietary rights to one's data may be abridged, and that the process of making one's data available to others will impose unacceptable time and resource costs upon the original investigators.

As to the concern that reanalysis may reveal mistakes in published work, or at least discrepancies about assumptions, appropriate procedures, and the meaning of findings, the results of past secondary analysis suggest such concern is warranted. Mistakes will be found, and genuine debates over procedure and interpretation will occur. However, such exchanges and the corrections that derive from them contribute to the advancement of scientific knowledge.

If researchers were required to make their data available for secondary analysis, they would be much more careful in the initial analyses and interpretations. Furthermore, the secondary analysts who publish "corrections" of others' work are themselves subject to the same process of reexamination and response and therefore must themselves be extraordinarily careful in their procedures and assertions. In the end, the public good and the cause of scientific exactness may be advanced, but the possibilities of egos being bruised in the process are sufficiently deterring that, Bryant and Wortman (1978) imply, unless professional associations and publication policies of journals force investigators to make their data avail-

able to others as a condition of publication, the data upon which most published work is based are likely to continue to be unavailable for secondary analysis and interpretation.

The matter of proprietary rights is much more complex. Certainly the original investigators have a prior claim to their own data. Generally investigators will not release data for public use until they have finished all the analysis they intend to do. If journals required that original data be made available as a condition of publication, perhaps those portions of data "released" in a given article might be made available for analysis by others prior to the release of the entire data set.

Some archives have handled the issue of proprietary rights by creating several categories of studies according to degree of accessibility to the general public, with data that are still being analyzed by original investigators sometimes available only by written permission of the investigators. Bryant and Wortman (1978: 384) suggest that negotiation may be necessary before potential analysts are allowed access to data that investigators have not finished using, with access contingent upon agreement on joint authorship or other rights such as "the privilege of published rebuttal and commentary."

The final issue on the part of the original investigators, cost, is more readily handled because there is successful precedent in the establishment and maintenance of many existing archives. Typically the user must pay the costs of duplication and preparation of material for secondary analysis.

Another potential problem of secondary analysis is that occasionally insufficient information is reported about the collection of the data to determine possible biases in them. A lack of specification about sampling, response rates, measurement and coding raises doubts about the quality of the data.

The use of data from a respectable source may hide incompetent analysis and interpretation by the secondary analyst. Inappropriate statistical tests may be used because the secondary analyst did not understand the data collection procedures sufficiently to know that they should not have been used and this error may pass unnoticed because of the quality of the data.

The largest U.S. archive of secondary data, the holdings of the Inter-university Consortium for Political and Social Research, deals with the problem of insufficient **documentation** by dividing their holdings into four classes according to the amount of effort devoted by ICPSR technicians and original investigators to standardizing and processing data sets to facilitate their use. The differences in the quality of documentation and preparation among studies archived by ICPSR is illustrated in this description of Class I and Class IV data sets:

Class I data sets have been checked, corrected if necessary, and formatted to ICPSR specifications. Also, the data may have been recorded and re-

organized in consultation with the investigator to maximize their utilization and accessibility. A codebook, often capable of being read by a computer, is available. This codebook fully documents the data and may include descriptive statistics such as frequencies or means. One copy of a printed codebook is supplied routinely to each Official Representative [of institutions belonging to the Consortium]. All Class I studies are available on magnetic tape in either card-image or OSIRIS [a computer software package] format. . . .

The Class IV studies are distributed in the form received by the ICPSR from the original investigator. Users of Class IV data should keep several considerations in mind.

Problems may exist which would not be known before processing begins, and thus the ICPSR can take no responsibility for the technical condition of the data. The requestor, therefore, must be prepared to accept some uncertainty as to the condition of the data. Requests for these studies will normally require a longer time to complete than more fully processed studies. . . . The documentation for Class IV studies is reproduced from the material originally received. . . . The majority of the studies in Class IV are available on magnetic tape. A few studies in this category are available only on cards because they contain multiple punches. (ICPSR, 1981-82: 19-20)

III. DOING SECONDARY ANALYSIS

Three key tasks in a secondary analysis are (1) designing the data collection, (2) identifying and coping with biases and errors in the selected data sets, and (3) choosing appropriate indicators and components of indexes in studies that were not designed to be combined or perhaps even to measure the concept that interests the secondary analyst.

For the secondary analyst the design of data collection and preparation consists of judicious selection from archival holdings or other repositories of those studies or portions of studies that are appropriate in time, location, and topic for the researcher's objectives.

A very popular resource for secondary analysts is the General Social Surveys (GSS), nationwide surveys of approximately 1,500 respondents that were conducted annually between 1972 and 1978 and at two-year intervals thereafter. The purpose of the GSS is to provide quality national data to the general social science community. The topics were chosen to meet the future interests of researchers and to provide data comparable to those gathered by polling organizations in earlier times. A core of items has been repeated in all of the surveys; therefore there are annual measurements of certain population characteristics and also data have been pooled so that the 1972-1980 cumulative codebook included frequency distributions for a national sample of more than 15,000 respondents.

The project is described by Glenn (1978: 532) as "the first sociological survey dealing largely with attitudes and behavior conducted for any purpose other than to provide data for research planned by the principal investigator or investigators." The response from the academic community

has been enthusiastic. Hundreds of articles and scores of monographs have been written using the GSS data as the major data base or for comparison with other data sets.

The data collection is done by personal interview by the National Opinion Research Center. The codebooks and data files are distributed at minimal cost by the Roper Public Opinion Research Center and the Inter-university Consortium for Political and Social Research. A summary of the limitations of the data, along with highly laudatory descriptions of benefits of the project for teachers, students, social researchers, and anyone else interested in societal characteristics, is found in a 1978 published symposium. Hyman's lyrical description of the surveys is an adequate introduction to the opportunities provided by this unique data set:

Springtime each year since 1972 brings the annual "General Social Survey" conducted by the National Opinion Research Center. . . . It does not, as the poet said, "droppeth as the gentle rain." It descendeth like a deluge, but it is, to use his words, "from heaven. It blesseth" the secondary analyst who appreciates and uses the lavish gifts showered on us all under the "National Data Program for the Social Sciences." . . .

The image that comes to mind as one browses through this codebook is the Christmas catalogue of one of those exclusive department stores catering to the rich—Neiman Marcus, for example. To be sure, the NORC items are not displayed in elegant fashion, but to the secondary analyst pure in heart they are as titillating as the mink parkas or "his and her airplanes" are to the Texas millionaire, and for social scientists they are just as valuable. The real difference is that one does not have to be rich to make a purchase. (Hyman, 1978: 545, 546)

Among the major contemporary archives of social science data are the Roper Center at Williams College and the Interuniversity Consortium headquartered at the University of Michigan. There are many other sources of secondary data. Hyman's (1972: 331–333) now-dated list of archives "functioning around 1970" included 17 institutions in the United States, Canada, England, the Netherlands, and Germany. Trochim's (1981) comprehensive essay on "Resources for Locating Public and Private Data" and Boruch, Wortman et al.'s (1981: 21–67) five-part section on "Policies of Key Agencies and Data Resources" provides easy-to-follow guidelines for gaining access to available data in private archives and the more complex procedures necessary for finding one's way among the mountains of federal data available for secondary analysis.

Let us illustrate the process with Trochim's (1981: 59–60) list of "guidelines of a search" and provide some insight into the quantity of data available by listing some of the catalogues of archives he describes. A secondary data search, Trochim says, consists of (a) *specification of needs* (examining subject indexes of archive holdings, identifying appropriate keywords); (b) *initial familiarization* (searching guides and catalogues and listing data ar-

chives or organizations that may have the data one wants); (c) *initial contacts* (contacting persons familiar with the archive in question and learning conditions of use); (d) *secondary contacts* (verifying details necessary to make formal requests); (e) *accessibility problems* (learning from people who have experience with using the data, and identifying probable difficulties); and finally (f) *analysis and supplemental analyses* (often as one does a secondary analysis additional data are needed, and the experience of going through the above steps facilitates repeating some of the process until the necessary data are acquired and analyzed).

The two main resources available to a potential secondary analyst searching for data are "printed catalogues, guides, and directories of archives and data bases," and certain organizations and user groups created to help researchers (Trochim, 1981: 60). Catalogues of archives include the *Encyclopedia of Information Systems and Services* (Kruzas, 1978), *Statistics Sources* (Wasserman and Paskar, 1977), the *Research Centers Directory* (Palmer, 1979), and the *Consultants and Consulting Organizations Directory* (Wasserman and McLean, 1976). The guides to federal data bases and available sources of governmental evaluation studies, longitudinal surveys, and demographic and economic surveys include *A Framework for Planning U.S. Federal Statistics* (U.S. Department of Commerce, Office of Federal Statistical Policy and Standards, 1978), *The Directory of Computerized Data Files and Related Software* (U.S. Department of Commerce, National Technical Information Service, 1978), the *Catalogue of Machine-Readable Records in the National Archives of the United States* (U.S. National Archives and Records Service, 1977), and *Federal Information Sources and Systems: A Directory for the Congress* (U.S. General Accounting Office, 1976), plus many catalogues that describe holdings of particular agencies or departments.

Private archives periodically print comprehensive catalogues of their holdings. As was already noted, a major private source is the Interuniversity Consortium for Political and Social Research, which publishes a yearly *Guide to Resources and Services* (1981-1982). Also a *Data File Directory* is published by the Association of Public Data Users (1977). Major organizations designed to serve secondary analysts include the Information Documentation Center of the National Technical Information Service, the Data Clearinghouse for the Social Sciences (Canada), the International Federation of Data Organizations for the Social Sciences, the European Association of Scientific Information Dissemination Centers, and the International Association for Social Sciences Information Service and Technology. Trochim's (1981: 65) Appendix lists mailing addresses for these and other organizations.

The data collection process ultimately comes down to securing catalogues or data lists from the archives, selecting studies that appear relevant, and ordering the codebooks to learn whether the variables of interest are available in usable form. If so, then the analyst purchases or borrows and

copies the data files. He or she also must decide whether to combine relevant portions of several data files into a master working file containing all the variables of interest—in effect, creating a new composite data file—or performing the various cross-tabulations and other analyses for each of the studies separately. Eventually, findings may be presented from several studies, each of which independently does or does not support a hypothesis.

There are many different ways to manipulate data files. These are really different designs of secondary analysis. Hyman (1972: 134–138) describes seven types of designs, which can also be combined in various ways.

1. The secondary analysis of a single survey
2. Pooling of multiple surveys, whereby many surveys that contain equivalent indicators are combined into a single huge data set
3. Internal replication with multiple surveys, in which “several previous surveys, all containing identical or equivalent indicators are incorporated within a single design” (Hyman, 1972: 134–135)
4. External replication by secondary analysis of a single survey, in which primary or secondary analysis of one survey is a replication of an earlier study which is either a primary or replicative survey, such that there is a linked time series of surveys
5. Intrasurvey replication, in which one survey has different indicators of the same variable, and each is analyzed separately
6. Truncating the sample, in which respondents not meeting certain criteria are excluded (“Rare” populations being studied by deleting all but the few possessing the rare characteristic is an example of extreme **truncation** of multiple surveys. See, for example, Reed (1976), in which 56 Gallup surveys were pooled to produce a sample of 85,957 respondents, 166 of whom possessed the desired characteristic, being Jewish and living in the South)
7. Synthesis of multiple surveys for broad characteristics of groups, in which narrowly defined characteristics are combined into a more broadly defined general category, thereby permitting analysis of the combined multiple surveys.

Much of Hyman’s (1972) book is a description of how varieties of these designs for secondary analysis can be or have been applied.

The second part of the design process is to deal with the problem of errors in the data, ranging from biases introduced in the initial data collection to errors in the coding and entering (or keypunching) stages of preparation for analysis by the original investigators. Errors in the original studies, perhaps properly allowed for there, or inconsequential to the original research objectives, may take on greater significance in a secondary analysis which has different objectives. The secondary researcher’s efforts to estimate the extent of error and to make allowances for it may be frustrated by poor documentation of the primary study, leaving the analyst with no means of assessing the nature and possible impact of procedures

not described; he or she must decide to use the data on faith (and here the reputation of the organization or individuals doing the original study is a critical element) or to exclude a study from reanalysis.

Finally, indicator selection and **index construction** are matters of concern because the secondary analyst must search among available indicators for those that seem to fit the conceptual framework. There is always the risk that the indicators were chosen because they "were there" rather than because they were indicators of choice; they may measure some concept other than that under study and therefore may introduce an unknown degree of bias.

The essential point is that combining and truncating, redesigning categories, abstracting general concepts from multiple indicators, and mixing and matching different data sets often require imagination and insight far beyond that needed to analyze a primary data set. Indeed, a problem facing social science in the current era of richness of available secondary data sources is the shortage of trained secondary analysts. One of the most successful secondary analysts in the United States, Norval Glenn, has summarized the "problem of underutilization" of data archives as primarily a problem of the creativity and technical skill of the present generation of researchers:

Highly successful users of archival data may always be rare, since good secondary analysis may require more imagination and ingenuity, and certainly it requires more tolerance for imperfection, than other kinds of research. However, if graduate programs in the social sciences were to make most new Ph.D.'s aware of the kinds of data available at the major archives, and if research design were no longer taught in such a way as to discourage the mode of thinking needed for secondary analysis, most persons regularly engaged in social scientific research would probably use archival data at least occasionally. And researchers would be less likely to waste money by gathering new data not distinctly superior to data in the archives. (Glenn, 1973: 44)

IV. ILLUSTRATIVE EXAMPLES

Secondary Analysis of Experimental Data: The Hawthorne Data

Nearly everyone who has taken an introductory social science course is familiar with the famous **Hawthorne effect**, reported to have occurred among workers at the Hawthorne plant of the Western Electric Company in Chicago who were studied in an extraordinary, complex series of experiments conducted between 1924 and 1932. The term "Hawthorne effect" refers to the reports by experimenters that workers subjected to variations in the lighting of their workplace (and other factors supposedly related to individual productivity) increased production regardless of the illumination

level. Expected declines in productivity associated with bad lighting did not occur. Eventually the researchers decided that because the research subjects had been "set apart" and felt that they were part of a special group, their morale and consequently their production increased.

The lengthy program of experiments on the relationship between various management styles and productivity eventually led to the conclusion that worker satisfaction was a critical variable intervening between contextual conditions, including management style, and output (Roethlisberger and Dickson, 1939). A direct consequence of the Hawthorne experiments and this interpretation of their findings was the "human relations" approach to management.

The Hawthorne experiments were conducted at a time when the systematic statistical analysis of social science data was in its infancy, and most of the researchers' conclusions were based on anecdotes, visual inspections and interpretations of rates of output as affected by variations in experimental conditions, and general summaries of the personal observations of the researchers. Perhaps their most influential findings were "that measured experimental variables had little effect, but that the unmeasured quality of human relations of workers to management and peer group was responsible for most output improvements observed in the first four experiments" (Franke and Kaul, 1978: 624).

Although the detailed numerical data from the eight-year program of research were retained in the Hawthorne plant, for 50 years there was no statistical analysis of the data. In 1976–1977 the data were borrowed from the Hawthorne plant (copies are now on microfilm in the libraries of the University of Wisconsin and Worcester Polytechnic Institute). Data from the first relay experiment, which had provided so much support for the "human relations" interpretation of the findings, were entered in computer files and analyzed by modern techniques of **regression** and **multiple regression**, by Richard Franke and James Kaul (1978). Fortunately, the data collection was superb for its time, with outputs graphed and most independent variables well described. Furthermore, enough is known about changes in certain historical factors (e.g., changes in employment opportunities between the boom times of the 1920s and the effects of the early Depression years on morale and work behavior) to treat them as variables in contemporary analysis.

Having obtained the "massive body of data and description," Franke and Kaul conducted a complex statistical analysis and demonstrated conclusively that the impressionistic findings that had spawned the human relations approach to business management were *not* supported by the Hawthorne first relay experiment. Instead, Franke and Kaul conclude:

... three variables—managerial discipline, the economic adversity of the depression, and time set aside for rest—explain most of the variance in quan-

tity of output for the group and generally for the individual workers. Two workers who exhibited undue independence from management . . . were replaced by two more agreeable workers. This exercise of managerial discipline seems to have been the major factor in increased rates of output for the now altered group, including increased productivity by the three individuals remaining. (Franke and Kaul, 1978: 636)

The findings of this first statistical interpretation of the Hawthorne studies are in direct and dramatic opposition to the findings for which the study is famous. "It was not 'release from oppressive supervision,' . . . but its reassertion that explains higher rates of production." The secondary analysts concluded that contemporary researchers' work should be directed more to the characteristics of *managers* and the processes by which they manage, and less to the human relations of the workers (Franke and Kaul, 1978: 636-638).

These spectacular findings, reversing truisms that had dominated management theory and distinguished one of the best-known social research projects ever done, were a direct result of painstaking secondary analysis made possible because the Hawthorne plant had retained the original data. The authors' final conclusion is a charge to modern social scientists to pay more attention to reexamining the research which forms the foundation of their disciplines:

The analytical procedures employed in the present study suggest feasibility of examining closely the building blocks of our disciplines, especially when quantitative information is available. This has long been done in the physical sciences, where development routinely includes the process of critical scientific review, secondary analysis, and replication of important studies. There appears a great need as well as opportunity for such activities in the social sciences. (Franke and Kaul, 1978: 638)

Secondary Analysis of Survey Data: Correlates of Widowhood Status

Harvey and Bahr (1974) were interested in learning whether there were differences between widowed and married persons in morale and in social involvement, particularly affiliation with organizations. They decided to use secondary data to test hypotheses derived from social-psychological and role-theory perspectives on the probable consequences of being widowed. They wrote to survey data archives for catalogues and scanned them to identify surveys which might include measures of morale, affiliation, and marital status in sufficient detail to distinguish widowed persons from the single, the divorced, and the married. It was also required that a potential data base for the study have a substantial number of the widowed, so that the effects of age and income could be separated from effects of widowhood. Codebooks were ordered for studies described as having detailed information on marital status as well as information on some of the

variables shown by prior researchers to influence the characteristics of the widowed.

Forty codebooks and sample descriptions were received and inspected, and this set of 40 yielded only three studies that had enough widowed respondents (operationalized at 150 or more) and sufficient data on other variables to justify further analysis. The three studies were (1) the Almond-Verba Political Attitudes Study, conducted in 1959–1960 in the United States, Italy, Germany, Mexico, and the United Kingdom; (2) the Bradburn Happiness Study, conducted in 1962 in four Illinois communities as a pilot study in psychological well-being; and (3) the Glock-Stark Religion in America Study, conducted in 1963 among respondents randomly drawn from membership lists of a sample of church congregations in four California counties. The Almond-Verba data included 4,892 respondents, of whom 476 were widowed; the Bradburn Happiness data was derived from 2,005 respondents, including 200 widows; and the Glock-Stark data set included 178 widows in a total sample of 2,871.

Each study had different indicators of morale, but there were comparable items on affiliation with formal organizations and on certain demographic characteristics including age and income. Analysis was done separately by study. The data sets were different enough that there was no attempt to pool samples into a composite data file. Data from the five nations represented in the Almond-Verba data were analyzed separately, in effect producing five different data sets. However, for all data sets the same comparisons between married and widowed people were possible, as were controls for respondents' age and income.

The basic findings drawn from the samples of respondents in five nations, including three very different regions in the United States, suggested that neither self-theory nor role-theory perspectives were adequate to explain differences between widowed and married persons in morale and affiliation. The subsamples of widows and the cross-cultural design provided, in effect, replications of this basic finding in seven different research settings. The results of these seven virtual replications by secondary analysis called into question some of the widely accepted generalizations about widows, many of them based on research among small, local samples. The authors' summary included a call for reexamination of data from previous studies that had not included adequate controls for income or poverty status:

... the negative impact sometimes attributed to widowhood derives not from widowhood status but rather from socioeconomic status. The widowed have appeared to have more negative attitudes than the married because they are much poorer than the married, and they have appeared less affiliated for the same reason. This does not mean that widowhood status has no effect upon morale or affiliation. . . . But the variation in these dependent variables attributable to widowhood status seems much smaller than that attributable to

income. Consequently, studies which have attributed long-term lack of affiliation or demoralizing consequences to widowhood should be reassessed to be sure that controls for income level have been adequate. There is need for more systematic documentation of the economic disadvantages of the widowed, and of the consequences of those disadvantages. The consequences of widowhood status *pe se*, more subtle and elusive, can only be highlighted when the powerful role of economic status has been taken into account. (Harvey and Bahr, 1974: 106)

Secondary Analysis of Evaluation Research: The Bureau of Indian Affairs Relocation Program

Much evaluation research is never released to the public but is retained by the sponsoring organization for use in modifying existing programs or justifying their continuation. However, the evaluations of many government programs, especially when the continuance of a program is at risk, are often highly controversial whatever the findings may be, and they stimulate controversy about the quality of the research procedures and the proper interpretation of the findings.

The evaluative work conducted as part of the ongoing management of agency programs is often ignored by the public and underutilized by its sponsor. This may be so because the agency needed only a small portion of the data collected, or because it was not satisfied with the findings and did not find them as immediately applicable as they had hoped, or simply because they lacked the technical expertise or the motivation to analyze the data beyond simple percentage distributions or tallies of the nature and quantity of services delivered.

An extensive follow-up of the Bureau of Indian Affairs' urban relocation program falls into this latter category. In 1966 BIA officials decided to evaluate the effectiveness of the relocation program by finding Indians who had been relocated from the reservations to cities in 1963 and given vocational training presumably suitable to their new environments. Respondents who had been relocated and trained three years earlier were now scattered across the nation, and substantial resources were spent finding them. Eventually 89 percent of a sample of 367 persons were located and interviewed by specially trained members of the BIA permanent staff. All interviewers and respondents were Indians. A superficial mimeographed report was prepared and then the data were put aside and forgotten.

Several years later two of the present authors learned about the project and, in conjunction with Lawrence Clinton, tried to locate the original data and obtain permission to analyze them. The major problem we faced was not that BIA officials were unwilling to have the data analyzed, but that no one knew what had happened to the interview schedules. Eventually they were located in a BIA administrator's office in a southwestern state, where they had been stored "temporarily" and then forgotten.

The original data were coded for computer analysis, and the existing research literature on characteristics of Indians who had been successfully relocated to urban areas and were steadily employed was surveyed for independent variables that had counterparts in the BIA follow-up study. A thorough statistical evaluation of the relative influence of relocation and BIA-provided vocational training revealed that (1) urban relocation, with or without vocational training, tended to improve the economic status of relocatees, but (2) the utility of the vocational training by itself was questionable at best. It had little measurable effect upon the subsequent employment records of relocatees (Clinton, Chadwick, and Bahr, 1975: 130-131).

By the time the secondary analysts located, analyzed, and published the results, the findings were essentially politically neutral, partly because they were dated. Nevertheless, we take the position that such an historical analysis of a program designed to help Indian people join the "economic mainstream" was worth doing because (1) there is very little published work on the effectiveness of programs designed to improve the employability of Indian people, and (2) the effectiveness of future programs aimed at ameliorating the poverty and unemployment of Indian people may be enhanced by the results of a competent evaluation of a previous program. The analysts were convinced that the data-collection procedures were carefully enough designed and implemented that the results would represent a fairly accurate assessment of the outcomes of relocation and employment training for the persons interviewed. Getting the results of the evaluation into the public record seemed an important initial step that would permit evaluators of similar programs to have some statistical benchmarks to compare with their own "success" rates.

Large-scale evaluations of controversial government policies and programs are likely candidates for secondary analysis because whatever their outcomes, there are powerful interest groups sufficiently biased either in favor of or against the program to justify independent analysis.

V. SUMMARY

Secondary analysis is the use of research materials for purposes different from the original research objectives, usually by persons other than those who collected the data. Secondary analysis is a method, not a classification of data quality as in the primary source-secondary source dichotomy used to classify documentary materials. Secondary analysis was used by prominent social scientists early in this century, but until recently it was not practiced extensively nor held in high repute as compared to primary data collection and analysis.

Secondary analysis offers substantial practical, social, and scientific-

theoretical benefits. The practical benefits include economies of time, money, and the good will of respondents. Secondary analysis provides more "findings per dollar" than any of the other research techniques described in this book. It saves time and points the way to successful avenues of inquiry.

As to the social benefits, it gives the sponsor/public/respondent more benefit per dollar or hour expended. Moreover, it provides freedom to do research free of bureaucratic constraints, to analyze creatively and follow the leads of one's own imagination, a luxury not always available to those who collected the original data.

Benefits for science and substantive knowledge include offering greater opportunity for comparative—cross-temporal, cross-cultural—research, making the study of trends and social change less costly, and providing an alternative to lengthy longitudinal research. Secondary analysis allows one to create retrospective or previous benchmarks, and these may be tied to future studies. Thus a scientist can "create a past" via secondary data, and measure present and create future measurements with his or her own primary data-collection efforts.

Another substantive benefit is that the techniques of pooling allow the creation of composite data sets large enough to create sizable subclasses of deviant or rare individuals, and thereby permit analysis about unique populations. Also there is the replication benefit: one can compare one's own results with those appearing in data collected by others or can use the prior studies as comparison points. Finally, secondary analysis stimulates theory-building and creative classification by requiring the combination of different but related variables. Thus it increases the likelihood that social science will be "self-correcting" in the sense that much secondary analysis involves some replication of previous findings as well as the discovery of previously unreported findings.

Another advantage of secondary analysis is that it permits specialization: persons uncomfortable with data collection can avoid it altogether, and those whose talents lie in index construction and analytical wizardry are freed from the necessity of collecting data. Data collection can be left to those whose interests and specialties are in sampling, interviewing, and the like.

Among the disadvantages of secondary analysis are the possible inapplicability of the data, for variables or populations studied may not be the ones the researcher needs; the data sets collected by others may be of low or inestimable quality because of poor documentation and coding; errors in the data may not be apparent to the secondary analyst; and data sets may not be available in usable form. Other problems are a prejudice against secondary analysis in some disciplines, as well as reluctance of investigators to make their data available, because they do not want to put "at risk" their own analysis and findings, or because the primary analysis has not been completed, or because the researchers do not want to bother

putting their data in archivable form. Problems of proprietary rights, protections of subjects' confidentiality, and costs of making data available to others in usable form are more easily handled and there already are well-established precedents.

Among the main stages of secondary analysis are the design of the data collection, identifying and coping with the biases and errors in data sets, and choosing appropriate indicators and components of indexes from studies that may not have been designed to yield adequate measures of the variables the analyst wishes to study.

The process of doing secondary analysis includes searching the catalogues of archives, identifying data sets or organizations likely to have the data one wants, making formal requests for the data, learning about the potentialities and problems of the data (preferably from others with experience in using similar data), and then performing the analysis according to accepted methods. Any resources can guide the search for secondary data, including printed catalogues, directories of archives and data bases, and organizations or user groups. Among the major archives presently used by social scientists are the Roper Center at Williams College and the Interuniversity Consortium headquartered at the University of Michigan.

There are several types of analytical designs for secondary analysis. The major obstacle to greater use of the massive data bases now available is the shortage of researchers trained in secondary analysis. There is still a bias in the training of social scientists such that primary data-collection techniques receive more attention than does secondary analysis.

Examples were given of secondary analysis of experimental data, of cross-cultural interview data, and of the findings of evaluation research. The secondary analysis of evaluation research is especially important because it provides independent confirmation of the findings of studies aimed at guiding policy decisions about large-scale programs that affect much of the population.

The secondary analysis of evaluation research often stimulates secondary critiques or reanalyses of the original data. Such iterations are valuable from the standpoint of the advance of scientific knowledge and the public good; data collected at great expense are more carefully assessed than they otherwise would be.

Often the benefits of secondary analysis far outweigh its costs. The availability of high-quality data sets is an extraordinary opportunity for creative investigators to look for meaningful findings relevant to the problems faced by today's societies and the theoretical perspectives that guide contemporary scientific work. The existence of the major archives makes high-quality data obtainable and comparatively inexpensive for serious, creative students at any level.

REFERENCES

- ASSOCIATION OF PUBLIC DATA USERS, 1977. *Data File Directory*. Arlington, Va.: Association of Public Data Users.
- BAILEY, KENNETH D., 1978. *Methods of Social Research*. New York: The Free Press.
- BORUCH, ROBERT F., PAUL M. WORTMAN, DAVID S. CORDRAY and ASSOCIATES, 1981. *Reanalyzing Program Evaluations*. San Francisco: Jossey-Bass.
- BORUCH, ROBERT F., DAVID S. CORDRAY, and PAUL M. WORTMAN, 1981. "Secondary Analysis: Why, How, and When." Pp. 1-20 in Robert F. Boruch et al., *Reanalyzing Program Evaluation*. San Francisco: Jossey-Bass.
- BRYANT, FRED B., and PAUL M. WORTMAN, 1978. "Secondary Analysis: The Case for Data Archives." *American Psychologist* 33 (April): 381-387.
- CLINTON, LAWRENCE, BRUCE A. CHADWICK, and HOWARD M. BAHR, 1975. "Urban Relocation Reconsidered: Antecedents of Employment Among Indian Males." *Rural Sociology* 40 (Summer): 117-133.
- FRANKE, RICHARD HERBERT, and JAMES D. KAUL, 1978. "The Hawthorne Experiments: First Statistical Interpretation." *American Sociological Review* 43 (October): 623-643.
- GLASER, BARNEY G., 1963. "Retreading Research Materials: The Use of Secondary Analysis by the Independent Researcher." *American Behavioral Scientist* (June): 11-14.
- 1962. "Secondary Analysis: A Strategy for the Use of Knowledge from Research Elsewhere." *Social Problems* 10 (Summer): 70-74.
- GLENN, NORVAL D., 1978. "The General Social Surveys: Editorial Introduction to a Symposium." *Contemporary Sociology* 7 (September): 532-534.
- 1973. "The Social Scientific Data Archives: The Problem of Underutilization." *The American Sociologist* 8 (February): 42-45.
- HARVEY, CAROL D., and HOWARD M. BAHR, 1974. "Widowhood, Morale, and Affiliation." *Journal of Marriage and the Family* 36 (February): 97-106.
- HYMAN, HERBERT H., 1978. "A Banquet for Secondary Analysis." *Contemporary Sociology* 7 (September): 545-549.
- 1972. *Secondary Analysis of Sample Surveys: Principles, Procedures, and Potentialities*. New York: John Wiley & Sons.
- ICPSR, 1981-82. *Guide to Resources and Services, 1981-82*. Ann Arbor, Mich.: Inter-university Consortium for Political and Social Research.
- KENDALL, PATRICIA L., and PAUL F. LAZARSFELD, 1950. "Problems of Survey Analysis." Pp. 133-196 in Robert K. Merton and Paul F. Lazarsfeld (eds.), *Continuities in Social Research: Studies in the Scope and Method of "The American Soldier"*. Glencoe, Ill.: The Free Press.
- KRUZAS, A. T. (ed.), 1978. *Encyclopedia of Information Systems and Services*, 3rd ed. Detroit: Gale Research.
- MERTON, ROBERT K., and PAUL F. LAZARSFELD (eds.), 1950. *Continuities in Social Research: Studies in the Scope and Method of "The American Soldier"*. Glencoe, Ill.: The Free Press.
- MOSER, C. A., and G. KALTON, 1972. *Survey Methods in Social Investigation*, 2nd ed. New York: Basic Books.
- NAROLL, RAOUL, 1962. *Data Quality Control—A New Research Technique*. New York: Free Press of Glencoe.
- PALMER, A. M. (ed.), 1979. *Research Centers Directory*, 6th ed. Detroit: Gale Research.

- REED, JOHN SHELTON, 1976. "Needles in Haystacks: Studying 'Rare' Populations by Secondary Analysis of National Sample Surveys." *Public Opinion Quarterly*, 40 (Fall): 514-522.
- ROETHLISBERGER, FRITZ, J., and WILLIAM J. DICKSON, 1939. *Management and the Worker*. Cambridge, Mass.: Harvard University Press.
- SIMON, JULIAN L., 1978. *Basic Research Methods in Social Science: The Art of Empirical Investigation*, 2nd ed. New York: Random House.
- STOUFFER, S. A., A. A. LUMSDAINE, M. H. LUMSDAINE, R. M. WILLIAMS, JR., M. B. SMITH, I. L. JANIS, S. A. STAR, and L. S. COTTRELL, JR., 1949. *The American Soldier: Combat and Its Aftermath*. Princeton, N.J.: Princeton University Press.
- STOUFFER, S. A., E. A. SUCHMAN, L. C. DEVINNEY, S. A. STAR, and R. M. WILLIAMS, JR., 1949. *The American Soldier: Adjustment During Army Life*. Princeton, N.J.: Princeton University Press.
- TROCHIM, WILLIAM M. K., 1981. "Resources for Locating Public and Private Data." Pp. 57-67 in Robert F. Boruch et al., *Reanalyzing Program Evaluations*. San Francisco: Jossey-Bass.
- U.S. Department of Commerce, National Technical Information Service, 1978. *Directory of Computerized Data Files and Related Software*. Washington, D.C.: U.S. Department of Commerce.
- U.S. Department of Commerce, Office of Federal Statistical Policy and Standards, 1978. *A Framework for Planning U.S. Federal Statistics*. Washington, D.C.: U.S. Department of Commerce.
- U.S. General Accounting Office, 1976. *Federal Information Sources and Systems: A Directory for the Congress*. Congressional Sourcebook Series. Washington, D.C.: U.S. General Accounting Office.
- U.S. National Archives and Records Service, 1977. *Catalogue of Machine-Readable Records in the National Archives of the United States*. Washington, D.C.: U.S. National Archives and Records Service.
- WASSERMAN, P., and J. MCLEAN (eds.), 1976. *Consultants and Consulting Organizations Directory*, 3rd ed. Detroit: Gale Research.
- WASSERMAN, P., and J. PASKAR (eds.), 1977. *Statistics Sources*, 3rd ed. Detroit: Gale Research.

Evaluation Research

elevan

I. Introduction

II. The Organizational Context of Evaluation Research

*The problem of organization
goals in evaluation*

Negotiating the evaluation

The politicization of evaluation

III. Evaluation Design

Design of program monitoring

Design of a program evaluation

IV. Evaluation Methods

Methods for program monitoring

observation/interviews

records and reports

surveys

Program evaluation methods

experimental design

quasi-experimental design

qualitative methods

V. An Example: The Juvenile Diversion Program Evaluation

The concept of juvenile diversion

Evaluation design

Evaluation results and conclusions

VI. Summary

I. INTRODUCTION

Evaluation research as it is now defined and practiced is of recent origin. It emerged from two related trends of the early 1960s. The first was the development and expansion of the political programs that were a part of the "Great Society" of the Lyndon Johnson administration. This administration operated under the philosophy that new social action programs could reduce much of the social inequality that existed in American society and correct many other social problems. The second trend was a new method of federal budgeting that emerged during this period called the Planning, Programming, and Budgeting System (PPBS). The central idea behind this system was that federal programs should be evaluated to see how well they were achieving their objectives. This information could then be fed back into the decision-making process (Williams and Evans, 1976: 293). Numerous federal, state, and local programs designed to cure poverty and eliminate ethnic and sex discrimination were thus created and many of these programs included a *mandated* program evaluation. The evaluation rush was on.

Evaluation research is generally considered to be applied rather than basic research. **Applied research** is oriented to finding solutions to specific social problems whereas **basic research** is typically directed toward advancing scientific knowledge for its own sake (Rossi, Wright, and Wright,

1978: 173). Some people see a sharp division between basic and applied research, but often the distinctions between them are in the researchers' attitudes and objectives rather than in the research activity *per se* (Rossi, Wright, and Wright, 1978: 172). At times the two may be indistinguishable (Rossi, 1980).

This chapter describing evaluation research will be somewhat unlike those that have gone before. Evaluation research is not *a* method. Rather, it applies a *variety* of techniques and methods to answer questions such as "Does a program work in the manner in which it was designed to work?" Thus, virtually all of the methods described previously and everything that we have learned about questions of design, measurement, and analysis will be used to conduct evaluation research. "What distinguishes evaluation research is not method or subject matter, but intent—the purpose for which it is done" (Weiss, 1972: 6). The primary intent of this chapter is to give an understanding of how to apply research methods already learned to the problem of program evaluation.

Evaluation research (or more simply, evaluation), then, is a general term that refers to a wide variety of research activities. Within the past 25 years evaluation research has come of age and there is now a vast published and unpublished literature on the topic. Several professional journals are devoted to evaluation, an annual review publishes the latest trends and research, and the study and practice of evaluation research have become a specialty in sociology, psychology, education, and economics. The amount of money spent on evaluation is staggering, perhaps over \$300 million in 1979 alone (Freeman and Solomon, 1981: 13). The topics covered by evaluation research are virtually unlimited. An examination of the titles in the *Evaluation Studies Review Annuals* reveals such topics as education, mental illness, public health, medicine, employment training, energy demand and conservation, auto repairs, and dog litter. In addition, there are articles on theory, methods, ethics, and utilization of evaluation. Indeed evaluation research spans the concerns of both scientists and the general public and attempts to bridge the gap between the two.

Despite all the activity, however, there is still disagreement about the definitions and purposes of evaluation research. Definitions abound but are not widely accepted because of the diversity of problems to which evaluation procedures are applied.

Rossi, Freeman, and Wright (1979: 32–51) identify four types of evaluation: program planning, program monitoring, impact assessment, and resource efficiency (Rossi, Freeman, and Wright call it "economic" efficiency). To illustrate these four types, let us suppose we want to set up a nutrition program for low-income women expecting their first child. In **program planning** our interest is in determining the target population (what constitutes "low income"), whether the program is needed, how many women are in the target population, what is the best way to reach

them, and what should be the information that the program gives to the new mothers. **Program monitoring** assures that the target population is being served, the information is being given out properly, the appropriate staff has been hired, and so on. Program implementation and monitoring focus on process. **Impact assessment** considers the products of the system. For example, with respect to the nutrition program questions would include whether the information was presented in an understandable way, whether the women changed their eating and health habits, and, if so, whether the changes were clearly attributable to the program. Finally, we may want to consider whether the program was the most cost-efficient way to get the information to expectant mothers. **Resource efficiency** issues relate to the possibility of equally effective but less expensive ways to achieve program objectives.

Each of these types of evaluation is important. Indeed, a comprehensive evaluation of a program would include all four. In practice, however, the most common types of evaluation are program monitoring and impact assessment. The principal reasons for the emphasis on these two types of evaluation are that (1) most programs are already in operation when an evaluation is started and therefore there is no opportunity for the evaluation of program planning, and (2) managers tend to be more concerned with the demonstrable effects (impacts) of a program than with the issues of whether some effects can be produced with lower resource inputs.

The present chapter treats these two topics of evaluation, program monitoring and impact assessment, in detail. In order to use common terms as well as to avoid confusion with social impact assessment (Chapter 12) we will use the term "program evaluation" rather than "impact assessment" when talking about the assessment of whether a program has achieved its objectives, and "evaluation research" will refer to the combination of program evaluation and program monitoring. The unit of analysis will be social action programs such as the nutrition program in the above example.

The key characteristics of applied or evaluation research programs are that they are client orientated, conducted within specific time frames, designed to provide information for policy decisions, and frequently developed to address research questions and variables that are specified by the client and so are beyond the researchers' control (see Rossi, Wright, and Wright, 1978; Rossi, 1980). However, perhaps the most distinguishing feature of evaluation research, as it pertains to social action programs, is that it is *organizational* research. Most social action programs are implemented by organizations such as schools, public health agencies, or local governments, and evaluation is conducted in and for these organizations. Evaluation cannot be separated from its organizational context; the nature of evaluation as well as the utilization of findings is determined by organizational factors and not the researchers. We will now discuss some of the more important of these organizational factors.

II. THE ORGANIZATIONAL CONTEXT OF EVALUATION RESEARCH

There is perhaps no form of research that is as constrained and influenced by its operating environment as evaluation research. Program evaluations are conducted in an organizational setting for organizational purposes. If the researcher can add theoretical or basic research questions to the evaluation, all the better, but the primary purpose of evaluation is to collect information about a procedure or program that may be useful in decision making by program managers.

Organizational factors determine much about how an evaluation is done. Cooperation of the organization can make an evaluation easier to conduct: it can open up files, provide access to employees, and so on. On the other hand, resistance can prevent the evaluator from getting data or entrance into the organization. Indeed, the evaluator may be caught between the demands of two organizations: the sponsor of the research and the implementing organization. For example, the Department of Education may wish to have a particular education program evaluated while a local school district may resist such evaluation.

There are several reasons why an organization or program staff may resist evaluation. Some common reasons given are fear that the program will be found ineffective, concern that evaluation will disturb the daily operation of the program, and the perception by staff personnel that the evaluation is of personnel rather than the program. In a juvenile diversion program evaluation (described later) some police officers resisted the study because they viewed it as reducing their discretion in how to handle a case. Generally resistance to evaluation can be overcome by careful attention by team members to the reasons for and sources of resistance.

Three other important organizational factors affect the evaluation of social programs. These are the problem of organization goals, the need for negotiated roles and research, and the politicalization of evaluation research. Each is discussed below.

The Problem of Organization Goals in Evaluation

One often-stated prerequisite for conducting a program evaluation is that there exist stated (or at least discoverable) program goals (Suchman, 1967; Nay et al., 1976; Rossi and Williams, 1972; Rossi, Freeman, and Wright, 1979). The purpose of evaluation, then, is to determine how well the program goals have been achieved. Two assumptions are implicit in the evaluation of programs with reference to their formal goals: (1) that organizations or social action programs have clearly stated goals against which their progress can be evaluated, and (2) that such goals do not change over the life of the program. Neither of these assumptions is necessarily correct. Uncritical acceptance of the "goal attainment" approach can lead

the evaluator into serious trouble, into what has been called the "goal trap" (Deutscher, 1976).

Organizations typically have two types of goals: official goals and operative goals (Perrow, 1961). **Official** (or sometimes called **public**) goals are those usually stated in official reports or in public statements by organizational leaders. They are also often used to justify the existence of an organization. **Operative goals**, on the other hand, are what the organization actually tries to achieve through the distribution of its resources or actions—what the organization actually does and accomplishes. Often there is little congruence between official and operative goals. The nutrition program described earlier provides an example. The stated, official goal of the program was to improve the nutrition of low-income women expecting their first child, but most of the program's resources were spent building political support or deciding what information to prepare for the potential clients.

The usual problem with official goals is that they are vaguely stated or too general; they are intangible goals. For example, some commonly expressed goals of social action programs are to "reduce poverty," "improve reading skills," "eliminate discrimination," or "make the streets safe for the public." It is impossible to describe the kinds of activities that can be used to achieve these intangible goals, and it is almost equally impossible to tell when the goal has been reached (Warner and Havens, 1968). Yet vague and intangible official goals are the rule rather than the exception—and because the social action programs of organizations reflect the official vague definition of ultimate objectives, the programs themselves are difficult to evaluate.

Since achieving such a general objective is virtually impossible, organizations do one of two things: either they substitute a more achievable goal (*goal diversion*), or they concentrate on program operations rather than on program goals (*goal displacement*) (Warner and Havens, 1968: 541). These substituted goals become the operative goals, toward which the resources of the program and activities of the staff are directed. There is also a third alternative, the substitution by program administration and staff of their personal goals for program goals, either official or operative. Accordingly, a major preliminary goal of a program evaluator is to discover the operative goals of the program, i.e., what are the workers in this program really trying to accomplish, in literal, measurable terms?

In addition to the typical discrepancies between official and operative goals, the evaluator must recognize that organizational goals change over time. New political attitudes may redirect a program, staff turnover may bring in new operative goals, or the organization may move through a series of operative goals seeking to find one with which "success" is possible. In sum, organizations and social programs are dynamic; change is the rule rather than the exception. Evaluators cannot assume that reports

of last year's or last month's objectives and procedures reflect the current reality.

There are two other reasons why evaluators cannot take the official goals of social programs at face value. First, the goals of the program (or even the program itself) may have been established for political reasons (because they are politically popular), not because there is any chance of their being achieved. Programs may be announced with great fanfare and publicity by officials to demonstrate how much is being done for "needy" populations (Weiss, 1975). Second, official goals for social programs are often established at the national level, yet local organizations set up to implement the program are geared toward local needs and preferences (Cook, 1981: 263) and thus serve local goals. The result is a discrepancy between the national and local goals, with the national goals going by the wayside.

In summary, in order to measure whether an organization or a social program has achieved its goals, the evaluator must answer several questions: Is the focus to be on the official or operative goals? Have the goals (official and/or operative) of the program changed, and if so, are the present goals or some previous goals to be used as standards in the evaluation? Moreover, since most programs have several goals, which goals have the highest priority?

The evaluator cannot answer these questions alone. General, specific, and operative goals must be identified by the sponsors of the evaluation, by the administrators of the organization or program, or by other interested parties, such as citizen action groups whose interests the programs supposedly serve.

Negotiating the Evaluation

Most evaluation research is conducted in and for organizations. Usually it is done to achieve organizational ends. Therefore, the evaluator rarely has total or even substantial control of the evaluation process (Suchman, 1967; Cronbach, 1982). He or she is not free to design and implement an evaluation without the approval and cooperation of many parties vitally interested in the evaluation's outcome. Thus the design and conduct of an evaluation are a negotiated process.

In fact, almost every aspect of the evaluation is open to negotiation with the relevant parties. It is essential that the evaluator negotiate *and record* a common understanding of the evaluation's objectives, methods, eventual audience, and potential utilization (Stufflebeam and Webster, 1981: 83). The evaluator, then, is part researcher, part administrator, but full-time diplomat. Furthermore, although it is essential that agreement be reached in advance about the matters mentioned above, the negotiation

process usually continues throughout the entire project. The major reason for the continuing communication and negotiation is that evaluation is not detached and objective, but highly charged, value-laden, and political in nature.

The Politicization of Evaluation

Evaluation research combines the potentially conflicting forces of sound research techniques, politics, and bureaucracy (Williams and Evans, 1976: 297; Weiss, 1975, 1976; Patton, 1978; Rossi, Freeman, and Wright, 1979). Weiss explains:

... evaluation is a rational enterprise that takes place in a political context. Political considerations intrude in three major ways, and the evaluator who fails to recognize their presence is in for a series of shocks and frustrations. First, the policies and programs with which evaluation deals are the creatures of political decisions. They were proposed, defined, debated, enacted, and funded through political processes; and in implementation they remain subject to pressures, both supportive and hostile, which arise out of the play of politics. Second, because evaluation is undertaken *in order to* feed into decision making, its reports enter the political arena. There evaluative evidence of program outcomes has to compete for attention with other factors that carry weight in the political process. Third, and perhaps least recognized, evaluation itself has a political stance. By its very nature, it makes implicit political statements about such issues as the problematic nature of some programs and the unassailability of others, the legitimacy of program goals, the legitimacy of program strategies, the utility of strategies of incremental reform, and even the appropriate role of the social scientist in policy and program formation. (Weiss, 1975: 13)

Thus, there are vested political interests in programs and in the outcomes of program evaluations. These vested interests include the program personnel, program backers and antagonists, funding agencies, clients, and various community interest groups. The evaluation of new programs is especially sensitive, for they may have not had time to establish firm political support for renewal funding, and consequently a negative evaluation report may signal the end of a program before it has had an opportunity to achieve its objectives. Indeed, the decision about which programs to evaluate is itself often a political decision (Weiss, 1975). Programs with lukewarm political support are more likely to be evaluated than those backed by powerful lobbies.

The politics of evaluation is not limited to vested political interests; the social science community also becomes involved:

Once evaluation studies are seen as likely to have important political consequences, they become fair game for people whose views are contradicted (or at least unsupported) by the data. A first line of attack is the study's methodology. Critics of every persuasion seem able to locate experts who

find flaws in the sampling, design, choice of statistics, measurement procedures, time span, and analytic techniques—even though their real criticisms derive less from methodology than from ideology. Whatever the motivation, a study whose conclusions enter the political arena must be prepared for searching scrutiny of its methods and techniques. (Weiss, 1976: 312)

Indeed, there are ways to show that *any* social program is ineffective. One list of such techniques (Gottfredson, 1979) includes these charges: the program was somehow contaminated; the wrong criteria were used to measure success; the past massive efforts have failed; the program is based on faulty theory; and finally, the fact that the program has been shown to work for some people does not mean it will work for everyone. The evaluator of a sensitive program can expect to be attacked from every quarter with a variety of reasons and techniques.

A “negative” evaluation does not automatically mean the demise of a program (Cook, 1981; Williams and Evans, 1976), any more than a “positive” evaluation guarantees its survival. Bureaucracies are amazingly resilient. A supportive political atmosphere is much more critical for a program’s continuance than the results of an evaluation (Weiss, 1975). An evaluation is usually *not* the determining factor in decisions about a program, but it does feed into the decision-making process.

A good example of all this is the Head Start Program, established in 1965. At the outset, it was to be a small experimental program reaching a limited number of children (Williams and Evans, 1976). The ostensible goal of the program was to help prepare disadvantaged preschool children for later school experience (McDill, McDill, and Sprehe, 1972). However, the idea was too good; it drew tremendous political support and was implemented nationally, involving nearly one-half million children and an annual budget of \$100 million (Williams and Evans, 1976). In June 1968 a comprehensive evaluation of the program was begun. The evaluation was an *ex post facto* study of children who had been in the Head Start Program and who, at the time of the study, were in the first, second, and third grades. These children were given a series of cognitive and affective tests and the results were compared with those of a constructed control group. The results of the study concluded that children who had attended Head Start were not appreciably different from the control group (Williams and Evans, 1976).

Immediately the evaluation was attacked from almost every side. Its methodology was attacked as weak at best; the tests used were criticized as inappropriate because Head Start had other goals than cognitive effects (e.g., health, nutrition, and community objectives); the tests themselves were said to be invalid; the sampling techniques were attacked as biased; and so on (Williams and Evans, 1976; McDill, McDill, and Sprehe, 1976). The political debate on the evaluation study was widespread and emotional. Almost nothing in the procedures or report escaped unscathed.

Nevertheless, the evaluation was rated technically as a fairly good one, easily the most comprehensive done up to that time.

The evaluation had little impact on Head Start. In 1968, prior to the evaluation, the budget for Head Start was \$323 million. After the study the budget for Head Start was \$360 million in fiscal year 1971 and \$376.8 million in fiscal year 1972 (McDill, McDill, and Sprehe, 1976: 166). It was the political process, and pressure, that saved the Head Start Program.

Evaluators are, to a degree, part of the political context within which their projects are enacted. Social programs are imbued with social and political values from a variety of sources, are enmeshed in conflicts of interest, and are influenced by various interest groups. To ignore the political context is to invite failure (Stufflebeam and Webster, 1981: 83). For some researchers, the political and bureaucratic infighting can be "good fun" (Rossi and Williams, 1972: xii). For those who prefer noncontroversial, noncommittal, or uneventful research, conducting an evaluation can be a maddening, frustrating experience.

In sum, the conduct of evaluation research is not merely a matter of theory and method. Indeed, the methods of evaluation may be of secondary importance; if the organizational and political components of an evaluation are not attended to, the technical competence of the researchers may be irrelevant. Thus, the design of an evaluation project is a balance between the organizational factors and the techniques of scientific research.

III. EVALUATION DESIGN

Designing an evaluation takes imagination, creativity, and a fair amount of political savvy. The research design specifies how resources will be allocated (Cronbach, 1982), and deciding how the limited resources will be used is the crucial decision. The design is the way the evaluator comes to know about the program and its consequences (Tharp and Gallimore, 1979), and it necessarily is tridimensional in time perspective: an evaluation is *of* the past, *in* the present, and *for* the future (Eraut, 1982).

Each evaluation research design is unique (Cronbach, 1982). There is no standard evaluation design because there are no fixed evaluation questions nor any set methods to be used in evaluation (Suchman, 1967; Rossi and Wright, 1977). The whole repertoire of research methods is available for use in an evaluation. Moreover, an evaluation design should be treated as a general guide, not as something set in concrete (Cook, 1981). Almost inevitably events occur that affect procedures or even prevent parts of the design from being conducted as planned. Accordingly, the design of an evaluation always includes contingency plans and fallback positions.

In designing an evaluation project, whether program monitoring or evaluation, the researcher must pay attention to the need for negotiation

with the interested parties. Program monitoring usually needs to be worked out with program administrators and staff, but program evaluation has a much wider audience.

Of course, there is no perfect evaluation design. There is a continuum of bad to excellent designs, and the complexity and dynamic nature of evaluation research may function to make "good" designs unworkable. The evaluator must recognize that he or she will not have full control over the research design and data analysis.

Having said that, let us turn to the design of both program monitoring and program evaluation. We will consider these separately, but in practice they are inseparable. To do a program evaluation without program monitoring leaves one in the position of not knowing how to interpret the results.

Design of Program Monitoring

As was stated earlier, social programs such as the nutrition program are organizational programs. Either an existing organization (the public health department) or a new organization—let's call it the New Mothers Nutrition Office—is established. A staff is hired, job descriptions are created, criteria are set up to determine how program clients are to be defined and served, and so on. Ultimately, the organization is responsible to provide services to clients. The purpose of program monitoring is to determine how the program operated and if the appropriate services were provided. Program monitoring is necessary because programs don't operate as designed or, as Patton (1978: 160) put it, "implementation of program ideals is neither automatic or certain." The establishment of a program is no guarantee that it performs the way it should.

There are two types of program monitoring. The first is an ongoing monitoring as the program is established and begins to deliver the services to clients. Such built-in monitoring is desirable but rare. Usually monitoring involves trying to reconstruct what the program has done after it has been discontinued. In these instances the evaluator turns historian and must "go back to the beginning" to discover what happened.

The first thing an evaluator needs is an understanding of the program. Why was the program established? What were its official and operative goals? Who was to benefit from the program? This information is basic for both program monitoring and program evaluation. One cannot evaluate a program he or she doesn't understand intimately.

When the objectives and history of the program are known, the next question is whether the program was implemented as intended. More than one evaluator has been embarrassed by evaluating program effects and pronouncing the results, only later to find the program was never implemented at all (see Patton, 1978: 149–150). More frequently programs are

only partially implemented. For example, the compensatory education programs called "Title I programs" (Hennessey, Takata, and Ames, 1980), and the juvenile diversion program described below are both instances of partial implementation.

When the evaluator is confident that implementation took place, the next step is to learn exactly what was done, and what portions of the original plans were deleted by conscious choice or default. Three different evaluations of program implementation should be done: effort, process, and treatment evaluation (Patton, 1978).

Effort evaluation is undertaken in order to assess the quantity and quality of activity. Effort questions are directed to finding out what staff and administrative positions were filled, whether the staff succeeded in getting program benefits to the intended clients, the qualifications of the staff, and so on. Programs that exhibit little activity or effort are not likely to be effective (Patton, 1978: 164–168).

The purpose of **process evaluation** is to examine the actual operation of the program, to see how the internal dynamics operate (Patton, 1978: 165). Among the important things to look for in process evaluation are the following: First, did the goals, either official or operative, remain the same, or were changes made? If changes were made in program goals or in administrative control, did they occur in the initial stages or after the program was established? These types of change have dramatic effects on program operations and, eventually, on project evaluation. The details about the day-to-day operation of the program include amount of staff turnover, decisions about unexpected events, resolution of internal conflicts, and staff morale. In short, process evaluation is based on a description of the strengths and weaknesses of the program (Patton, 1978).

Process evaluation is developmental, descriptive, flexible, and inductive (Patton, 1978: 185). It may be used to help improve program performance, feeding back information to program administrators that assists their decision making about the operation and course of the program. The main use of process evaluation is in interpreting the results of program evaluation. To know that a program is effective or ineffective is not enough. The evaluator must be able to explain why a given level of effectiveness persists, and much of that "why" is imbedded in details about program operation. Expansion of programs or even their continuance at a given scale is facilitated by the information on effective process or system weaknesses identified in process evaluation.

Most social programs provide some type of service or treatment for a selected population of potential clients. Moreover, it is anticipated that the service provided has some intended, positive effect on the target population. For example, a nutrition-education program rests on the assumption that knowing more about nutrition will improve the health of low-income mothers and their babies. The "treatments" may be classes in

nutrition, pamphlets on good nutrition, radio or TV announcements, or some combination of these. That part of evaluation research known as **treatment evaluation** is intended to document exactly what treatments were given, usually in the context of what treatments were intended to be given and why they were presumed to be effective.

Treatment evaluation may include some additional considerations. First, was a treatment administered at all? Was it the "correct" treatment? Was the treatment standardized, or was it administered haphazardly or influenced by personal characteristics of the staff? (Rossi, Freeman, and Wright, 1979: 132-133). It is essential that the evaluator know the kind of treatment given, and its mode of administration, or program evaluation is impossible. As an illustration, let us return to our nutrition program. Assume that the treatment intended was the provision of classes in nutrition for expectant mothers. Assume also that classes were never actually conducted; instead, pamphlets were distributed. An evaluation is called for, and the hired evaluator, accepting the program description that the treatment consisted of nutrition classes, proceeds to assess the effects of the classes. At some point the researcher will learn that the "treatment" being evaluated is not the treatment that occurred. Thus, treatment evaluation is closely related to the evaluation of degree of program implementation.

One other element needs to be considered in program monitoring, that of assessing the characteristics of the client population (Rossi, Freeman, and Wright, 1979). Of primary interest here are the traits of the population receiving treatment, how they were selected (by volunteering, referrals, or random assignment), and the characteristics of those who dropped out of the program or refused the service. A detailed understanding of the target population and the characteristics of that portion served is essential if results are to be generalized to other populations and treatments.

It cannot be expected that programs, even under the best of conditions, will perform exactly as designed. There will always be some difference between the real and the ideal. Given this reality, Patton (1978: 158-160) has raised the troubling but crucial question of how to decide when a program has been implemented "enough" so that it can be evaluated meaningfully. That is, how close does program implementation in practice need to be to the implementation planned in the program design? Decisions about acceptance thresholds of implementation must be jointly made by the evaluator and the program administrators. If the sponsor of the evaluation or the program administrators believe that implementation was "good enough" to merit assessment of program consequences, the evaluator can generally proceed, provided that the level of implementation is clearly documented as part of the research process.

Only when the evaluator has a thorough understanding of the program, its purposes, operation, and degree of implementation, can attention fruitfully be turned to assessing the affects of the program on its clients.

Design of a Program Evaluation

The objective of program evaluation is to establish whether a social program is producing its intended effects (Rossi, Freeman, and Wright, 1979: 161). In terms of our illustrative nutrition program, the objective would be to find out whether the program made any difference in the nutrition of the target population. A related purpose of program evaluation is to explain why a program did or did not produce the intended effects.

Technically, program evaluation involves selecting independent and dependent variables and creating effective indicators of these variables. Thus, all of the problems of measurement discussed in earlier chapters apply to program evaluation, as do the concerns about internal, external, and construct validity that apply to experimental research. In these regards, evaluation does not differ from social science research generally. The same standards of good research apply.

The first stage of design is to arrive at the *evaluation question*. The evaluation question is simply: *What information is wanted from the evaluation? What research questions about the effects of the program are to be answered by the results of the evaluation?* All other decisions are made with respect to these central questions. One may arrive at the evaluation question in different ways. Sometimes the evaluation question, along with the independent and dependent variables, are "givens" defined by administrators or program sponsors at the outset (Suchman, 1967). The evaluator may have some input in designing measurement of variables, but often his or her influence on the central evaluation questions is limited to such technical input.

In situations where the evaluator has more input, Patton (1978) recommends that the key evaluation questions be worked out between the evaluator and relevant decision-maker. Patton defines a relevant decision-maker as one who will use the evaluative information to make decisions about the program and thereby limits the people making input to those linked to the program by formal organizational ties. This may be too limited a set of "advisors" especially in contexts of high political sensitivity. Evaluations may have many different audiences, and if only to minimize later headaches, the evaluator may wish to consider key evaluation questions as perceived by other interest groups. The opportunity to do so is contingent upon the willingness of those who are funding the evaluation to allow the evaluator such freedom.

Once the evaluation question has been formulated, a second, equally important decision must be made: *What will constitute a successful result* (Suchman, 1967: 109)? That is, what standards will be used to assess actual program effectiveness? No program is fully effective, and any program can be shown to be relatively unsuccessful. We have already noted the problem with using official goals to define "success," and even operative goals are rarely fully implemented. The dilemma is how to choose meaningful stand-

ards for evaluation for programs known to be only partly effective when, at the same time, program results must be judged by some objective, external criteria. There is no definitive solution to this dilemma. One possible approach derived from organizational theory is that standards be selected which independent observers would agree represent "satisfactory" though not optimal performance (Scott, 1977: 68-69; Suchman, 1967). For example, one could use an "expected" subjective level of performance, or select an acceptable percentage of the potential client population as a standard of satisfactory achievement. Such a negotiated establishment of realistic standards of success usually proves more useful than would evaluation tied to some lofty official or ideal operative goals.

Rarely is there only one evaluative question. Programs have multiple goals, many of which may need to be evaluated. For example, evaluation questions appropriate to our illustrative nutrition program might include: Did expecting mothers change eating and health habits? Which had the greater effect on behavior, the classes or the pamphlets? Among women who did change their eating and health habits, how was the baby's health affected? When there are multiple evaluation questions, there must be some way to rank their importance, for resources are never adequate to deal with all questions on an equal basis.

Another important issue is that of the timing of program effects. It is always possible for an administrator to say, "The program is working well but it is too early for us to see the results." The illustrative nutrition program is designed to have immediate effects; the women's nutrition and their health presumably should improve while they are pregnant. On the other hand, the effects of a reading program may not show up for several years. The credibility of an evaluation may depend on establishing *in advance* whether measurable effects should be anticipated immediately, after some specified period of intermediate length, or only after many years or even generations.

The anticipated point at which the positive consequences of a program should appear is basic to the decision about when to evaluate. It makes little sense to conduct an evaluation of program effectiveness early in a program's existence if it is designed to produce effects only in the long run. However, it might make sense to do program monitoring during the inception and early history of a program to make sure that implementation is occurring as planned. Therefore, it is important to match the timing of the evaluation with the timing of the intended effects. In programs whose effects are to be long-term, evaluation becomes expensive in the sense that clients must be followed up years after their participation, and at the same time changes in program definition and administration over the years must be charted.

Two other aspects of social programs need to be taken into account in designing program evaluations: expected effects and unintended effects.

We noted earlier that some programs are established for political reasons regardless of the evidence as to whether or not they will have demonstrable positive effects. However, almost any program has some consequences. Even programs created by well-intentioned people and designed to provide real assistance to needy target populations are often designed by people who have little expertise in the creation of effective organizations (Chen and Rossi, 1981). Often such programs reflect the "amateur" perceptions of a few powerful people rather than sound organizational policy. To correct for this condition, or rather permit evaluation in spite of it, Chen and Rossi (1981) suggest a multigoal theory-driven approach to evaluation. Its basic notion is that for a particular program treatment and population, it is possible to estimate, from social science theory or past program experience, some expected effects of the program. This approach serves two purposes. First, it can take the design of program evaluation away from an exclusive emphasis on program goals, either official or unofficial. Second, it increases the probability that *some* of the effects of the program, unanticipated or intended, will be identified. Other approaches, especially those focusing on explicit goals, have found most programs to have "no effects," a finding which, understandably, has been a continuing disturbance to sponsors and administrators of social programs.

Related to this point is the brute fact that many programs do not have positive effects, and most produce both positive and negative results. Program actions have negative as well as unintended effects. In fact, one systems analyst argues persuasively that programs are more likely to have negative than positive effects and that the probabilities that a new program will improve things are quite low. Among his principles relevant here are: "Systems in general work poorly or not at all"; "New systems generate new problems"; "Complicated systems produce unexpected outcomes"; "The system itself does not do what it says it is doing"; and "In complex systems, malfunction and even total nonfunction may not be detectable for long periods, if ever" (Gall, 1975: 125-130).

It is not possible to formulate research designs guaranteed to discover unintended effects, but evaluators must watch for them. What is needed, in addition to perceptivity, sensitivity, and imagination, is a consciously designed approach to identifying possible program effects and careful analysis of the data as they are collected.

IV. EVALUATION METHODS

Evaluation design is the setting forth of the questions that need to be answered by the evaluation, along with the plan for collecting and analyzing the data. The nature of the design helps determine appropriate ways to evaluate the program. In this section we consider possible ways to

evaluate programs. These are separated into methods for program monitoring and methods for program evaluation.

Methods for Program Monitoring

Programs are monitored to learn how they are operating. The methods used for program monitoring include (1) observation/interviews, (2) records and reports, and (3) surveys.

Observation/Interviews There is no substitute for personal observation. Nothing gives the evaluator a better "feel" for program operation than watching firsthand the daily functioning of each aspect or stage of a program. It is necessary to observe operations at different times in a given day, on different days, and at different points in the monthly or annual "calendar of operations" to ensure that all relevant states/contexts of operation are observed. If the evaluator is fortunate enough to be involved at the outset of a program, then he or she should observe, and perhaps even participate in, the planning sessions in which programs and services are designed. Staff meetings, the processing of clients by the program personnel, interpersonal relations of staff members, and the physical design and ecology of program operation all merit systematic attention.

Observation does not have to be passive. Evaluators should talk informally with program administrators, program staff, clients, and others in the community interested in or connected with the program. More formal interviews may also be conducted to ask specific questions about program operation, to assess people's perceptions about program goals, and to determine the receptiveness of potential clientele to the program and its aims. The observations and interviews serve to "ground" program operation in an "objective" (i.e., not bound by program policy or image) way that is not obtainable from program records and reports.

Records and reports Organizations maintain records and make reports for reasons of fiscal accounting, management requirements, budgeting needs, and public relations. These written accounts, while biased or incomplete in many ways, may contain useful data about program operations.

Client records and administrative files are generally available. Client records may reveal client characteristics, types of service rendered, dates of service, types of "termination" of service, amount of service rendered, and so on. Administrative records presumably would yield information on number of clients processed, staff turnover, cost figures by category permitting the calculation of cost-effectiveness ratios, memos that indicate staff problems and how they were resolved, overall impressions of the prevailing management style and its consequences, changes in operative objectives,

and many other qualitative dimensions of program operation. Not all organizations keep accurate records, however, nor the same kind of records. Moreover, because the records are maintained for administrative reasons, they may not contain the kind or quality of information needed for research. Furthermore, there is always the question of accuracy versus expediency in the initial collection or subsequent compilation of official records.

Records and reports can be used to make a program look good, to enhance its importance, or to hide its inadequacies. An observation by Sir Josiah Stamp aptly makes this point:

The government are very keen on amassing statistics. They collect them, raise them to the n th power, take the cube root and prepare wonderful diagrams. But you must never forget that every one of those figures comes in the first instance from the village watchman, who just puts down what he damn pleases. (quoted from Nettler, 1974: 45)

Surveys Surveys, either by questionnaire or interview, are useful in estimating the level of services actually received by clients, as well as people's attitudes about programs generally or the quality of service obtained. They are also useful in assessing community support for programs and determining the size of the potential clientele. Periodic surveys can be an essential part of program monitoring and can provide reliable estimates of trends in the community in which a program functions as well as in the specific outcomes of the program itself.

Program Evaluation Methods

The central problem of evaluation design is the selection of data-collection techniques. There is no particular method that constitutes evaluation research, and different methods are appropriate for different types of research questions. There are no absolutely correct choices, nor any completely wrong ones. Rather, there are varying degrees of appropriateness, and the training of the investigator is often the critical factor. As in other aspects of evaluation, there are some general considerations to be weighed before one makes the final decision about the research method.

The program organization is a factor in the decision. The method of preference may be appropriate for the evaluation question but unusable because of organizational constraints. For example, a randomized experiment may be the best method to test a new way of teaching college algebra, but if the program has been in effect for a year when the evaluator is called in, there is no longer the possibility of random assignment of students to treatment and control groups. Similarly, participant-observation may be the method of choice to evaluate a group counseling technique. However, adding a new member to an existing group might be disruptive to other

group members, or it might change the nature of the group sufficiently to bias any findings. It is also possible that program administrators and/or evaluation sponsors may simply veto a particular method. Their reasons might range from real concern with program operation to personal preference. Thus, even the choice of method is a negotiated issue in evaluation research. Some methods simply are not appropriate in particular organizations, and trying to "force" a method may lead to virtually useless findings or to unnecessary conflict and resistance from those whose goodwill is essential to the completion of the project and the utilization of its findings.

One other problem to be considered in the selection of research methods is the problem of detecting "small effects." One of the most common findings of evaluation research is that programs have no effects on client populations (Chen and Rossi, 1981; Bennett and Lumsdaine, 1975; Gilbert, Light, and Mosteller, 1975). While there are many reasons for this, an important one is that short-run effects of most interventions tend to be small (Rossi and Wright, 1977; Gilbert, Light, and Mosteller, 1975; Sechrest and Yeaton, 1982). A program's effects, though positive, typically appear in minor increments which may take considerable time to accumulate in significant improvements. Of course, slow, steady improvement is not what is promised by the proponents of many "quick fix" programs. In choosing methods of data collection, the researcher will normally want to select approaches sensitive to small program effects.

Furthermore, the evaluator will have to interpret the meaning of the small effects observed. For example, when the television program *Sesame Street* was evaluated, it was found that preschool children who watched the program could identify two more letters of the alphabet than those who did not watch the show (Rossi and Wright, 1977: 18). While this result was statistically significant, the difference is certainly not an earthshaking one; the ability of certain preschoolers to identify two more letters of the alphabet than others is not of substantive significance. In assessing the meaning of such small effects, consider the distinction made by Sechrest and Yeaton (1982), who ask: Given a small effect, what would be the total effect if that small difference were aggregated over a large population. Some effects, like that of *Sesame Street*, if aggregated over all preschool children, still would not be important because the letter difference is not additive. That is, 50,000 children who watched *Sesame Street* would not know 100,000 more letters of the alphabet than 50,000 who did not. Thus, the benefits of watching *Sesame Street* had implications at the individual level only. On the other hand, if a family planning program leads all women of child-bearing age to reduce their total fertility by an average of .25 children, then the aggregate social effects are sizable and have major societal implications. Thus, the "quality" of the quantitative change, as well as its size, is important.

These two examples are clear-cut and the substantive differences are obvious. In most situations, however, the meaning of small differences is not so obvious. As Sechrest and Yeaton suggest:

We do not, unfortunately, know much about assessing quality of change. At present we can do little more than urge investigators to consider carefully the quality of the changes they are investigating and to attempt to use their quality estimates to form their judgments about the importance of the . . . changes they are able to demonstrate. (Sechrest and Yeaton, 1982: 159)

In selecting methods for conducting program evaluation, then, the evaluator must realize that most programs show only small effects on the target population, and that some effects have only an individual importance whereas other effects have general social significance. The choice of method must keep these potential problems of interpretation in mind. The goal of program evaluation is to be able to make some intelligible statement about the program, preferably one that has implications for program modification and improvement.

Having reviewed these issues, it is now possible to consider the major methods of program evaluation: (1) experimental design, (2) quasi-experimental design, and (3) qualitative methods. Rather than an extensive discussion of these methods, some of the arguments for and against their use in evaluation will be presented. These are not the only methods that are used in program evaluation, but they are the ones used most frequently.

Experimental design Since the recent growth of interest in evaluation, experimental design has been held up by many as the ideal, or even the only method of program evaluation. The reasons for the stature of the experimental design is that only through true randomization of subjects into treatment and control (or alternative) groups—which controls for all extraneous variance—can one measure the effects of the program treatment with assurance. The goal of experimental design is prediction and control: prediction of effects to other groups or situations, and control of variance.

Many programs have been evaluated by experiment. Gilbert, Light, and Mosteller (1975) describe several program evaluations using this design, and the "juvenile diversion" evaluation presented later in this chapter was evaluated by an experimental design. Yet, when compared with the total volume of evaluation research, true experimental designs are uncommon. There are several reasons for this. First, many program evaluations are commissioned long after the program has started, making it impossible to randomly assign program clients to different groups.

Even when an experimental design is possible, there is often sufficient resistance by administrators or program sponsors to prevent it. Much of this resistance centers around the assignment of clients to treatment or

control groups by random choice. A common argument against randomization is that withholding of treatment from some of the target population is unethical, unfair, or illegal (Cook, Cox, and Mark, 1977; Rossi, Freeman, and Wright, 1979). This argument is made even if the proposed randomization is to alternative treatments rather than treatment and control (no-treatment) groups.

Other problems arise from the fact that experimental design is often impractical for the evaluators. First, it is very expensive (Rossi, Freeman, and Wright, 1979). Both the treatment and control (or alternative treatment) groups must be pretested, both must be carefully monitored over the duration of the program, and then both must be tested at the conclusion. Not all program evaluations are funded well enough to permit such expensive evaluation. Second, experimental designs take a long time to complete (Rossi and Wright, 1977), sometimes as long as several years to conduct a true experiment, analyze the results, and write and present the report. Program administrators and funding agencies want an evaluation done quickly, perhaps before the next budget proposals are due.

Finally, program organization is a confusing factor. One assumption of experimental design is that there is a clearly defined and manipulated independent variable that in turn has a causal effect on a dependent variable. This assumption does not fare well in program organization. The "treatment" often is not clearly enough defined to be interpreted as a single variable, it may not be consistent throughout the program, and it may not even fit the category of an independent variable but instead be more appropriately seen as an intervening or one of a combination of variables, most of which are unmeasured and uncontrolled (Guttentag, 1977).

The justifications for using an experimental design are based in the theory, methodology, and philosophy of science. The arguments against using experimental designs are based in the practicalities of organizations and politics. Experimental design is probably most effective when a program is new or when the treatment is something that can be easily and completely standardized, such as transfer payments (Rossi and Wright, 1977). Evaluators should not automatically rule out experimental design because of the difficulties surrounding it any more than they should not believe that experimental design is the only way to do program evaluation.

Quasi-experimental design Quasi-experimental designs are generally thought of as a step down from experimental design; they are second best. The use of such methods is generally advocated when true experimentation is not possible. The question of how effective they are is open. On one side is the belief that, if properly conducted, quasi-experimental designs are as effective as experimental designs (Rossi, Freeman, and Wright, 1979). The argument against using quasi-experimental designs is that unless the evaluator has controlled all the relevant variables, there are plausible al-

ternative explanations for the outcomes (Campbell and Boruch, 1975: 202). That is, quasi-experimental designs leave open the question of whether variation in the dependent variable is due to the program treatment or to other variables.

Perhaps the most commonly used quasi-experimental design is that of the constructed control group. To illustrate this, let us return to our nutrition program example. Suppose that it was considered unethical to withhold this program from any eligible women, so an experimental design was not used. An alternative way to estimate the effects of the program would involve "constructing" a control group suitable for comparison with the women served by the nutrition program. The needed control group can be created in several ways: one good method would be to interview women who had a first birth during the time of the program. Because the nutrition program was directed at low-income women expecting their first child, the control group would have to be limited to low-income women. It might also be desirable for the control group to be like the nutrition program clients in age, race, and education. Obviously, the larger number of variables we want the two groups matched on, the more difficult it will be to construct a control group, and the larger will have to be the initial group (or universe) of women who had a first birth during the period the nutrition program was in operation. When the control group has been constructed, both groups are compared in dietary habits, daily nutrition, or whatever indicators of program "success" are being used. If the program group has better dietary habits and so on, then we may wish to credit some of that positive difference to the program. If there is no significant difference, we may try to explain that in another way, (e.g., the nutrition program has not been in operation long enough for differences to appear). If there is a significant difference between the two groups, the evaluator must decide whether it is due to the program, to other factors, or to both.

There is always the problem of selection bias. Even if the two groups are similar on all the characteristics thought relevant, it may be that women who came to the nutrition program were more interested in nutrition than the general population represented in the constructed control group. They may already have had significantly better nutrition habits than did women in the control group. Thus, there is at least one plausible explanation other than that the nutrition program made a difference. There are likely to be others. Such possible alternative explanations are always present in quasi-experimental design.

It is generally thought that quasi-experiments are to be used when true experiments are not possible. Other reasons for using quasi-experiments are that they are much less expensive and quicker to do than experiments, and these are important factors in program evaluation.

Qualitative methods Both experimental and quasi-experimental methods are based on quantitative measurement of variables. Although it

is widely believed that quantitative measurements are required for objective, reliable, and valid research, some evaluation researchers (Deutscher and Gold, 1979; Patton, 1978, 1980) question the need for quantitative measurement. They argue that qualitative methods not only are as useful as quantitative ones, but may in fact be more realistic in program evaluation. Deutscher and Gold say that programs with vague or intangible goals can and should be evaluated with qualitative methods. In contrast, researchers who prefer quantitative data claim that only programs with clear quantitative goals can be evaluated at all (Nay et al., 1976).

Patton asserts that there is no qualitative "method." Rather, one deals systematically with whatever kinds of data or information are collected (Patton, 1980: 22). Qualitative data, he writes,

... consist of quotations from people and descriptions of situations, events, interactions, and activities. The purpose of the data is to understand the point of view and experiences of other persons. (Patton, 1980: 36)

Whereas the explicit purpose of quantitative research is prediction and control, the purpose of research using qualitative methods is said to be understanding the meaning of events to people (Patton, 1980: 45).

Qualitative data usually derive from field methods that are inductive in nature. Inferences are made about the effectiveness of a program on the basis of reports of program clients, program staff, and other informants, rather than pretest and posttest measurements. Thus, a program, while it may not show a quantitative improvement or a statistically significant result, can be shown to be meaningful and important to the clients. To return to the nutrition program, the experiences that expectant mothers get in the nutrition program, including such indirect benefits as being able to make new friends, may have great meaning to them, even though quantitative changes in dietary habits may be miniscule. Moreover, by getting "close" to the program (Patton, 1980), an evaluator who uses qualitative methods may be in a better position to say why a program succeeded or failed. The use of qualitative methods requires that the evaluator be personally involved in more aspects of the program under scrutiny.

All in all, qualitative methods seem more likely to reflect how program participants react to a program and how they feel they were affected by it. This dimension of program impact may not be measured quantitatively. On the other hand, the personal identification with or positive feelings about the worth of a program's indirect benefits may be irrelevant to the stated objectives of a program.

In summary, an evaluator should never think strictly in terms of one method, nor in terms of quantitative versus qualitative approaches. The best evaluations use multiple methods, including combinations of quantitative and qualitative approaches. As we have argued throughout this book, multiple approaches to assessing research problems are usually preferable to a single method.

V. AN EXAMPLE: THE JUVENILE DIVERSION PROGRAM EVALUATION

The Concept of Juvenile Diversion

In 1976 an evaluation was conducted to determine how effective different juvenile diversion strategies were in reducing recidivism among juvenile offenders. "Diversion programs" were set up to divert juvenile offenders away from the juvenile justice system, that is, to keep them away from juvenile courts and probation departments.

The idea of diverting young offenders from full exposure to the juvenile justice system was stimulated by the President's Commission on Law Enforcement and the Administration of Justice, and such programs became part of the "national strategy" for delinquency (Klein, 1976). Diversion programs enjoyed widespread political support and monetary backing.

The idea of diversion programs was grounded in some widely accepted sociological thinking. The philosophical basis was a set of theoretical notions generally referred to as labeling theory (Hooper, Smith, and Bohnstedt, 1976). This perspective assumes that official labeling (i.e., such formal "rites of passage" as court convictions) of youths as "delinquent" sets them apart and stigmatizes them. Official labeling is also thought to lead to "secondary deviance" by limiting the individual's access to conventional roles, and incarceration is seen as impelling "contaminating" interaction with "hard core" delinquents. The specific assumptions, or model, underlying diversion programs were:

1. Diversion programs will prevent youth from entering the juvenile justice system—or at least reduce their level of penetration.
2. Diversion programs will cost less than traditional justice processing.
3. Diversion programs will be less stigmatizing than traditional justice system processing.
4. Diversion programs will reduce "contaminating" contact with other juvenile offenders.
5. Diversion programs will reduce recidivism among diverted youths. (Smith, Bohnstedt, and Tompkins, 1979)

Despite widespread political and financial support and a substantial theoretical basis, there was virtually no empirical evidence that diversion programs reduced recidivism. There were several reasons for this lack, but perhaps most notable were inadequate methods of evaluation. One of these evaluation problems included the lack of a clear-cut definition of diversion (juveniles least likely to repeat their deviant acts were most likely to be placed in diversion programs). In addition, many of the programs were demonstration projects, with all the problems inherent in new organiza-

tions. Finally, even in programs where juveniles were to be randomly placed in diversion programs, true random assignment was not accomplished.

Evaluation Design

In such an environment a new evaluation of juvenile diversion programs was conducted (Hooper, Smith, and Bohnstedt, 1976). The diversion program was implemented by a large municipal police department in southern California. Researchers working with the California Youth Authority (CYA) approached the police department and asked if the police were interested in a randomized experiment to evaluate the effectiveness of several diversion strategies. The police department agreed to participate in the study.

The primary intent of the evaluation study was to determine, by a carefully controlled experimental design, whether juvenile offenders randomly assigned to alternative diversion strategies were as likely to be rearrested as were offenders in a control group which received no diversion services. The measure of diversion effectiveness was recidivism rates six and twelve months after arrest. If a juvenile had not been rearrested by that time, diversion was considered a success. Diversion into an alternative program occurred following an arrest but before further action was taken. Specifically, it directed the juvenile offender away from court or probation. The control group consisted of offenders who were counseled and released. They received no further services.

Two groups of youthful offenders were included in the study. The first group of "less serious offenders" (including those arrested for offenses such as loitering and being drunk and disorderly) consisted of youths who normally would have been diverted under the existing program (the police department had an ongoing diversion program, which is one reason why it was asked to participate in the study). The second group of "more serious offenders" (including offenses for things like burglary, assault with a deadly weapon, sexual assault, and extortion) normally would have been referred to probation with a request for probation supervision (referred to as a non-detained [NDP] request). This inclusion of more serious offenders was one of the strong, and new, aspects of this evaluation of diversion programs.

The research was done in three phases. The first phase was an experimental design in which eligible subjects were randomly assigned to one of two possible diversion strategies. Those who normally would have been sent to the ongoing police department diversion program were randomly assigned to one of three treatments: counsel and release (by the police department); in-house diversion counseling (the ongoing program within the police department); or referral to a community resource agency.

These last two treatments constituted the diversion strategies. Those who normally would have been subject to an NDP were randomly assigned to one of four treatments: the three already mentioned or an NDP.

This experimental design phase called for each juvenile officer to routinely process the juvenile offender and determine the normal disposition for the person, either diversion or an NDP. Once this was done, the officer called the CYA office, identified himself (by photo number), and told the CYA operator what the normal disposition would be. In return, the CYA operator, using a random device, would tell the juvenile office which disposition to use. This strategy took the randomization procedure out of the hands of the police officer (having the officer determine the random disposition had been a source of bias in previous evaluations).

Prior to implementation, the research design for the first phase of the evaluation was presented to police department officials, both in general departmental administration and in the juvenile division. Some objections and concerns were raised which were settled by negotiation between the research staff and police department officials. First, it was agreed that despite the need for randomized dispositions, concerns about labeling and stigma dictated that no juvenile should receive a more serious treatment than he or she normally would have received. This provision had two implications for the research. First, cases that were simple "counsel and release" were excluded from the study, and second, juveniles who would have been sent to diversion under normal circumstances were not subject to a nondetained petition. Also, dependency cases were not included in any part of the evaluation.

The evaluation staff was also cognizant of considerations for community safety on the part of the police department. While the study was designed to include the more serious offenders, juveniles who were the subject of detained petitions were excluded from the randomized disposition process because they were regarded as threats either to public safety or to themselves. A later decision was made to exclude all juveniles already on active probation in order to preserve the continuity of the working relationship between the police and probation department.

The second phase of the evaluation consisted of follow-up field interviews with all subjects who could be located and would consent to be interviewed. The interview schedule was designed to measure family and peer interaction patterns, self-conceptions, school adjustment, self-reported delinquency, and the juveniles' perception of whether diversion services had influenced his or her behavior.

The final phase of the study involved determining recidivism rates of the randomly assigned groups. This was done by checking police department records for rearrests at both six and twelve months after the initial arrest. At the same time, data were gathered from juvenile files on family,

on demographic and offense factors believed to be predictive of recidivism, and on the diversion services received by the juvenile.

Evaluation Results and Conclusions

A total of 512 juveniles were randomly assigned to the four different dispositions (Smith, Bohnstedt, and Tompkins, 1979). The distribution of the cases is shown in Illustration 11.1.

As can be seen, there were few cases in the NDP category; this is typical, for relatively few arrested juveniles need an NDP under normal conditions. After exclusion of cases for various reasons (juveniles who turned 18 during the study period, who moved, who were not city residents, etc.) there were 390 cases for which rearrest data were available. There were no significant differences between the groups on demographic and offense characteristics thought to be predictive of recidivism.

How effective were the different diversion strategies in reducing recidivism? Results of the evaluation suggest that the diversion procedures made little or no difference in recidivism. At the end of the six- and twelve-month intervals after initial arrest there were no significant differences between the groups in percent rearrested (see Illustration 11.1).

Among the less serious offenders, those who were simply counseled and released were no more likely to be rearrested six or twelve months

ILLUSTRATION 11.1 **Recidivism rates of youthful offenders by random disposition.**

TYPE OF DISPOSITION IF NOT RANDOMIZED	COUNSEL AND RELEASE	IN-HOUSE DIVERSION	COMMUNITY DIVERSION	NDP REQUEST	ALL DIS- POSITIONS
Less Serious Cases (Diversion Pool)	N = 95	N = 105	N = 88	—	N = 288
Rearrested at six months	31%	23%	31%		28%
Rearrested at twelve months	39%	31%	42%		37%
More Serious Cases (NDP Pool)	N = 29	N = 22	N = 29	N = 26	N = 106
Rearrested at six months	35%	41%	41%	35%	38%
Rearrested at twelve months	45%	59%	48%	62%	53%

Overall recidivism rate at six months was 30%, at twelve months, 41%.

Source: Smith, Bohnstedt, and Tompkins (1979).

later than those who were assigned to one of the two diversion strategies. Among the more serious offenders, those assigned to the NDP did not have a significantly higher rearrest rate than those assigned to counsel and release or to one of the diversion programs. In sum, the diversion programs did not seem to affect recidivism.

The data gathered in the second phase field interviews shed some light on these findings. A total of 180 usable interviews were conducted. Of those juveniles assigned to the community diversion treatment only 41 percent actually received agency services. Almost half (48 percent) of the interviewed community diversion subjects said that they had not been referred to a community agency, and another 11 percent said that they had been referred but did not go to the agency.

In the in-house diversion disposition, juveniles were referred to a police department diversion officer or a student intern working under the direction of an officer. Only about half (49 percent) of those assigned to in-house diversion said they ever returned to the police department after their initial release. Unfortunately it is not possible to say why those juveniles who did return to the police department did so. Some may have returned for further investigation and final disposition (e.g., those arrested on the weekends or at night), and others may have returned for diversion counseling. Some may have returned for both reasons.

However, those who did receive diversion services, either by a community agency or by the police department, had recidivism rates about the same as those who received no additional services. Indeed, some fragmentary data point to a pattern that those who received no additional services had lower recidivism rates than those who received diversion services.

In summary, no definitive statement could be made about the effectiveness of juvenile diversion programs from this evaluation. The principal reason for this is that the diversion services were not really provided, i.e., the program was not fully implemented. Although juveniles were referred to diversion programs, most of them did not receive additional services. This clearly demonstrates that although randomization may be successful from a technical standpoint, organizational operation as well as individual actions may subvert a well-designed study, making conclusions about a nonimplemented program impossible.

Returning to the five assumptions concerning juvenile diversion programs, the evaluation staff concluded:

1. Diversion programs may bring juveniles further into the juvenile justice system. Many who in other circumstances would be counseled and released were now sent to diversion programs, thus "widening the net" of the justice system.
2. Diversion programs are more costly than traditional programs, because diversion programs for youth who normally would have been counseled and released require additional, and probably unnecessary, expenditures.

3. The third assumption can be challenged by the observation that most juveniles sent to diversion would not normally have been processed further into the justice system. Community diversion most frequently involved referral to a clinical casework agency, treatment that might involve the stigma of psychiatric maladjustment. The in-house division required return visits to a police facility or other city-owned neighborhood facility. It seems reasonable to suppose that such repeat contacts may have carried more stigma than outright counsel and release, the disposition that most diverted youths apparently would have received had the diversion program not existed.
4. With respect to whether diversion programs lead to less "contamination" with other delinquents, there is some evidence to suggest that certain group programs in the inner city may have increased contact with a peer group of juvenile offenders. There is need for further research in this area, but the assumption that diversion reduces "contaminating" contact with other delinquents is at least questionable.
5. The findings reported above do not support the assumption that diversion reduces recidivism.

Finally, the evaluation staff made two recommendations. First, that diversion programs concentrate on a higher-risk juvenile offender than has traditionally been served. Low-risk offenders (first-time offenders, status offenders, and those committing minor misdemeanors) could be counseled and released while those who normally would have been processed beyond the point of being subject to an NDP could be diverted. Second, where diversion is granted, monitoring procedures should be developed to assure prompt disposition and service provision. It is essential that the program be implemented, that is, that diversion services actually be provided.

Shortly after this evaluation study was completed, the funds for diversion programs were sharply reduced and the program evaluated was eliminated. Public and political opinion was shifting from a philosophy of protecting juveniles from the harshness of the judicial system to one of punishment and deterrence.

VI. SUMMARY

Evaluation research is organizational research. It is done in organizational settings and for organizational purposes. This fact places several constraints on evaluation research as compared to other research. The evaluator has much less control over the research in terms of selecting the relevant variables, the time element and the population to be researched. Even the basic research question may be partly or entirely out of an evaluator's control.

Political forces are also involved in evaluation research. Conducting an evaluation of a social program brings to the fore questions about program philosophy, how well programs are meeting their objectives, or if they are needed at all. The results of evaluation necessarily are interpreted and used

in a political context and for political ends because programs are politically supported and opposed, and because programs must justify their existence through the political processes of budgeting and legislation.

There are two major types of evaluation: program monitoring and program evaluation. Program monitoring is done in order to gauge how well a program is operating, to determine the official (those stated in official reports or public statements by organizational leaders) and operative (what the organization actually tries to achieve) goals, and to identify strengths and weaknesses of the program as it was implemented. Program evaluation assesses whether the program had any effect on its target population. Most effects identified will be relatively small and will emerge gradually over time. Rarely does a social program have an immediate, obvious effect in the direction anticipated by program goals.

There are many methods available for program monitoring and program evaluation. The whole range of social science methodology is employable in evaluation research. There is no one correct method for doing an evaluation. The methods of choice must be those that best fit the evaluation needs and are compatible with the program organization. As with other research, multiple methods are vastly superior to monomethod approaches.

Until recently, experimental design was regarded as the best way to evaluate a social program. Nowadays other methods are recognized as appropriate under certain conditions. Qualitative methods are advocated along with alternative quantitative modes of evaluation. The choice of evaluation methods should not be seen as an "either-or" choice between quantitative and qualitative approaches, but rather as combining both types of research to yield as richly varied and complete a picture as possible of the program's activities, benefits, and costs.

REFERENCES

- BENNETT, CARL A., and ARTHUR A. LUMSDAINE, 1975. "Social Program Evaluation: Definitions and Issues." Pp. 1-38 in Carl A. Bennett and Arthur A. Lumsdaine (eds.), *Evaluation and Experiment*. New York: Academic Press.
- CAMPBELL, DONALD T., and ROBERT E. BORUCH, 1975. "Making the Case for Randomized Assignment to Treatments by Considering the Alternatives: Six Ways in Which Quasi-Experimental Evaluations in Compensatory Education Tend to Underestimate Effects." Pp. 195-296 in Carl A. Bennett and Arthur A. Lumsdaine (eds.), *Evaluation and Experiment*. New York: Academic Press.
- CHEN, HUEY-TSYH, and PETER H. ROSSI, 1981. "The Multi-Goal, Theory-Driven Approach to Evaluation: A Model Linking Basic and Applied Social Science." Pp. 38-54 in Howard E. Freeman and Marian A. Solomon (eds.), *Evaluation Studies Review Annual*, Vol. 6. Beverly Hills, Cal.: Sage.
- COOK, THOMAS D., 1981. "Dilemmas in Evaluation of Social Programs." Pp.

- 257-287 in Marilyn B. Brewer and Barry E. Collins (eds.), *Scientific Inquiry and the Social Sciences*. San Francisco: Jossey-Bass.
- COOK, THOMAS D., FAY LOMAS COX, and MELVIN M. MARK, 1977. "Randomized and Quasi-Experimental Designs in Evaluation Research: An Introduction." Pp. 103-135 in Leonard Rutman (ed.), *Evaluation Research Methods: A Basic Guide*. Beverly Hills, Cal.: Sage.
- CRONBACH, LEE J., 1982. *Designing Evaluations of Educational and Social Programs*. San Francisco: Jossey-Bass.
- DEUTSCHER, IRWIN, 1976. "Toward Avoiding the Goal-Trap in Evaluation Research." Pp. 249-268 in Clark Abt (ed.), *The Evaluation of Social Programs*. Beverly Hills, Cal.: Sage.
- DEUTSCHER, IRWIN, and MARGARET GOLD, 1979. "Traditions and Rules as Obstructions to Useful Program Evaluation." Pp. 107-140 in Norman K. Denzin (ed.), *Studies in Symbolic Interaction*. Greenwich, Conn.: Jai Press.
- ERAUT, MICHAEL, 1982. "Handling Value Issues." Pp. 271-286 in Ernest R. House, Sandra Mathison, James A. Pearsol, and Hallie Preskill (eds.), *Evaluation Studies Review Annual*, Vol. 7. Beverly Hills, Cal.: Sage.
- FREEMAN, HOWARD E., and MARION A. SOLOMON, 1981. "The Next Decade in Evaluation Research." Pp. 12-26 in Robert A. Levine, Marian Solomon, Gerd-Michael Hellstern, and Hellmet Wollmann (eds.), *Evaluation Research and Practice*. Beverly Hills, Cal.: Sage.
- GALL, JOHN, 1975. *Systematics: How Systems Work and Especially How They Fail*. New York: Pocket Books.
- GILBERT, JOHN P., RICHARD J. LIGHT, and FREDERICK MOSTELLER, 1975. "Assessing Social Innovations: An Empirical Base for Policy." Pp. 39-193 in Carl A. Bennett and Arthur A. Lumsdaine (eds.), *Evaluation and Experiment*. New York: Academic Press.
- GOTTFREDSON, MICHAEL R., 1979. "Treatment Destruction Techniques." Pp. 110-126 in Lee Sechrest, Stephen G. West, Melinda A. Phillips, Robin Redner, and William Yeaton (eds.), *Evaluation Studies Review Annual*, Vol 4. Beverly Hills, Cal.: Sage.
- GUTTENTAG, MARCIA, 1977. "Evaluation and Society." Pp. 52-62 in Marcia Guttentag and Shalom Saar (eds.), *Evaluation Studies Review Annual*, Vol 2. Beverly Hills, Cal.: Sage.
- HENNESSEY, MICHAEL, GLENN TAKATA, and NANCY L. AMES, 1980. "Title I Selection Methods and the Coverage of Disadvantaged Students." *Evaluation Review* 4 (February): 43-57.
- HOOPER, DOUGLAS A., PEGGY SMITH, and MARVIN BOHNSTEDT, 1976. "Evaluation of Juvenile Diversion." Department of Sociology, California State University, Long Beach. Unpublished paper.
- KLEIN, MALCOLM, et al., 1976. "The Explosion in Police Diversion Programs: Evaluating the Structural Dimensions of a Social Fad." Pp. 101-119 in Malcolm Klein (ed.), *The Juvenile Justice System*. Beverly Hills, Cal.: Sage.
- MCDILL, EDWARD L., MARY S. MCDILL, and J. TIMOTHY SPREHE, 1972. "Evaluation in Practice: Compensatory Education." Pp. 141-185 in Peter H. Rossi and Walter Williams (eds.), *Evaluating Social Programs*. New York: Academic Press.
- NAY, JOE E., JOHN W. SCANLON, RICHARD E. SCHMIDT, and JOSEPH WHOLEY, 1976. "If You Don't Care Where You Get To, Then It Doesn't Matter Which Way You Go." Pp. 97-120 in Clark Abt (ed.), *The Evaluation of Social Programs*. Beverly Hills, Cal.: Sage.
- NETTLER, GWYNN, 1974. *Explaining Crime*. New York: McGraw-Hill.
- PATTON, MICHAEL QUINN, 1978. *Utilization-Focused Evaluation*. Beverly Hills, Cal.: Sage.

- . 1980. *Qualitative Evaluation Methods*. Beverly Hills, Cal.: Sage.
- PERROW, CHARLES, 1961. "The Analysis of Goals in Complex Organizations." *American Sociological Review* 26 (December): 854-866.
- ROSSI, PETER H., 1980. "The Challenge and Opportunities of Applied Social Research." *American Sociological Review* 45 (December): 889-904.
- ROSSI, PETER H., HOWARD W. FREEMAN, and SONIA R. WRIGHT, 1979. *Evaluation: A Systematic Approach*. Beverly Hills, Cal.: Sage.
- ROSSI, PETER H., and WALTER WILLIAMS (eds.), 1972. *Evaluating Social Programs*. New York: Seminar Press.
- ROSSI, PETER H., JAMES D. WRIGHT, and SONIA R. WRIGHT, 1978. "The Theory and Practice of Applied Social Research." *Evaluation Quarterly* 2 (May): 171-191.
- ROSSI, PETER H., and SONIA R. WRIGHT, 1977. "Evaluation Research: An Assessment of Theory, Practice, and Politics." *Evaluation Quarterly* 1 (February): 5-52.
- SCOTT, W. RICHARD, 1977. "Effectiveness of Organizational Effectiveness Studies." Pp. 63-95 in Paul S. Goodman and Johannes M. Pennings (eds.), *New Perspectives on Organizational Effectiveness*. San Francisco: Jossey-Bass.
- SECHREST, LEE, and WILLIAM H. YEATON, 1982. "Assessing the Effectiveness of Social Programs: Methodological and Conceptual Issues." Pp. 151-166 in Ernest R. House, Sandra Mathison, James A. Pearsol, and Hallie Preskill (eds.), *Evaluation Studies Review Annual*, Vol. 7. Beverly Hills, Cal.: Sage.
- SMITH, PEGGY, MARVIN BOHNSTEDT, and TRACY TOMPKINS, 1979. "Juvenile Division Evaluation: Report of an Experimental Study." Paper presented at the Annual Meeting of the Pacific Sociological Association, Anaheim, Cal.
- STUFFLEBEAM, DANIEL L., and WILLIAM J. WEBSTER, 1981. "An Analysis of Alternative Approaches to Evaluation." Pp. 70-85 in Howard E. Freeman and Marian A. Solomon (eds.), *Evaluation Studies Review Annual*, Vol. 6. Beverly Hills, Cal.: Sage.
- SUCHMAN, EDWARD A., 1967. *Evaluative Research*. New York: Russell Sage Foundation.
- THARP, ROLAND G., and RONALD GALLIMORE, 1979. "The Ecology of Program and Evaluation: A Model of Evaluation Succession." Pp. 39-60 in Lee Sechrest, Stephen G. West, Melinda A. Phillips, Robin Redner, and William Yeaton (eds.), *Evaluation Studies Review Annual*, Vol. 4. Beverly Hills, Cal.: Sage.
- WARNER, W. KEITH, and A. EUGENE HAVENS, 1968. "Goal Displacement and the Intangibility of Organization Goals." *Administrative Science Quarterly* 12 (March): 539-555.
- WEISS, CAROL H., 1972. *Evaluation Research: Methods for Assessing Program Effectiveness*. Englewood Cliffs, N.J.: Prentice-Hall.
- . 1975. "Evaluation Research in the Political Context." Pp. 13-26 in Elmer L. Struening and Marcia Guttentag (eds.), *Handbook of Evaluation Research*, Vol. 1. Beverly Hills, Cal.: Sage.
- . 1976. "The Politicization of Evaluation Research." Pp. 301-322 in James E. Anderson (ed.), *Cases in Public Policy Making*. New York: Praeger.
- WILLIAMS, WALTER, and JOHN W. EVANS, 1976. "The Politics of Evaluation: The Case of Head Start." Pp. 242-309 in James E. Anderson (ed.), *Cases in Public Policy Making*. New York: Praeger.

Social Impact Assessment

twelve

I. Introduction

II. Why Do Social Impact Assessment

Federal regulatory agencies

State and local government concerns

Industry concerns

III. What Is Included in a Social Impact Assessment

Factors affecting the nature and magnitude of social impacts

IV. Doing Social Impact Assessments

Defining the task

baseline profile

baseline projections

project description

impact projections

impact analysis

A triangulated approach

using secondary data

using survey methods

using participant observation

The importance of a triangulated approach

V. Impact Mitigation and Monitoring

VI. Summary

I. INTRODUCTION

Social Impact Assessment (SIA) is a comparatively new application of social research that has developed largely as a result of legislative mandate. However, SIA should not be viewed as a "new" or even a special type of research method. In fact, many of the issues discussed in this chapter have been treated in previous chapters. Our primary concern here is to discuss how the methods of survey research, participant observation, and secondary data analysis are applied in studying the community changes associated with large-scale energy resource development, industrial development, or other actions that change existing social and ecological patterns.

To provide a brief example of the application of social impact assessment methodology, we can cite a study recently completed by the authors dealing with some of the potential consequences of designating large tracts of federally managed lands in western states as "wilderness areas." When a section of land is designated as a wilderness, no one is allowed to live on it permanently, no motorized vehicle travel is allowed, and an effort is made to preserve the land in a condition as near as possible to its primitive state.

The Federal Bureau of Land Management (BLM) controls 23 million acres in the state of Utah, which is 40 percent of the total land area. In some counties the BLM manages 95 percent of the land. In 1976 Congress

passed the Federal Land Policy and Management Act, which initiated a review of public lands for possible wilderness designation. Nearly 5 million acres in Utah were targeted for study. Designation as wilderness has very profound consequences on the region as well as on the local community. Gas and oil exploration, coal and uranium mining, the extraction of oil from shale and tar sands, the grazing of cattle and sheep, and the harvesting of timber are severely limited. At present private citizens can, for a fee, graze livestock, explore for energy, or harvest timber on federal land. To limit these activities has rather significant impacts on development and growth in an area.

Information was collected from government and private agencies about population, number of head of livestock, amount of minerals extracted, and so on in the area (secondary analysis). Local newspapers were content-analyzed, especially letters to the editor, for clues about public attitudes toward designating a particular tract as wilderness. In addition, community leaders, public officials, and civic, religious, and business leaders were interviewed about their perceptions of what the impacts in the area would be. Obviously the responses were mixed, as cattlemen, sheepmen, timbermen, and miners were bitterly opposed to "locking up the land" while environmentalists and those involved in recreational activities, especially backpacking and river running, were very much in favor of "preserving the natural state." Community leaders tended to oppose wilderness because to so designate large tracts of nearby land, in their view, would probably limit population growth and economic development. The social impact assessment did not resolve these conflicts of interest but did clearly identify them and provided public officials with valuable information to guide decisions about wilderness.

Social impact assessment merits a separate chapter for two reasons: First, while SIA is not a unique methodology, it is a unique application/combination of other methodologies to a particular type of problem; second, many students of social science will be involved in this type of research in the future. Few social scientists were doing this type of work a decade ago, but now many employees of universities, governments, and consulting firms do social impact assessments. It therefore seems important that students have at least a rudimentary understanding of SIA.

Typical problems that require social impact assessment include large-scale energy development projects such as the construction of a coal-burning power plant, the opening of a new strip-mine, the building of a nuclear power plant, and the construction of a hydroelectric dam. Other types of projects include creating a nuclear waste storage facility, expanding a weapons assembly plant, constructing a highway, building a power transmission line, building a water storage and canal system for agricultural irrigation, deploying a weapons system such as the MX missile, and so on. In each of these cases, the proposed project would probably increase

employment opportunities in the selected areas and produce demographic and economic growth. It could also change dramatically the composition of the local population, their preferred lifestyle, or the balance of political control. Extensive planning and preparation are necessary if serious disruptions of the existing communities are to be avoided or at least minimized. Administrators and planners, in cooperation with social scientists, are finding it important to assess the "social impacts"—the potential social, economic, demographic, public service, and fiscal changes that will derive from a proposed development (Leistritz and Murdock, 1981: xiii).

Impact is defined as changes in current or established economic, demographic, and social conditions that are caused by the introduction of a project such as those described above. The nature and magnitude of the impact vary from project to project but are most directly affected by the combination of two sets of factors: characteristics of the project itself and characteristics of the impact region or community and its inhabitants.

Early efforts at identifying "social impacts" were often journalistic, painting a largely negative picture of the declining quality of life in modern "boom towns." Typical headlines were: "Crime Increasing Sharply in West's Boom Towns," and "Trailer Houses, Row on Row—In Most You'll Find Woe on Woe." The communities that sprang up were often referred to as "aluminum ghettos" and were characterized as being inhabited by people afflicted with the "three D's"—divorce, drunkenness, and depression. Illustration 12.1 describes the events assumed to be typical of rapid growth communities.

We do not discount all of these early descriptions. In some impacted communities crime, family disturbances, delinquency and truancy, marital discord, drinking and drug abuse, and mental health problems did increase significantly with the immigration of many unattached male construction workers (see Weisz, 1979; Pattinson, Weisz, and Hickman, 1979; Gilmore and Duff, 1975; Lovejoy, 1977; McKeown and Lantz, 1977; Lantz and McKeown, 1979; Freudenburg, 1979, 1982; Moen et al., 1981).

In addition to potential negative impacts, the social impact researcher must assess the positive changes in current conditions that will occur with the introduction of some type of development. Whether increases in major social pathologies will occur, or whether the net consequences of the project will be positive, are empirical questions that must be answered anew for each project.

II. WHY DO SOCIAL IMPACT ASSESSMENT

There are at least three sets of factors responsible for the current emphasis on social impact assessment. These include new federal rules and regulations, state and local government concerns, and industry concerns.

ILLUSTRATION 12.1 *SOME EARLY DEPICTIONS OF IMPACTS
ASSOCIATED WITH RAPID GROWTH*

GILLETTE, Wyo.—The little girl was about four, with blonde hair and large, sad eyes. She sat quietly and said nothing, while a nurse cleaned her infected ears and changed the filthy clothes she was wearing.

"When we found her she was curled up in a baby's crib in a trailer about five miles out of town," said Robert Weisz, a psychologist and director of the Northern Wyoming Mental Health Center here. The girl's mother was having a nervous breakdown and her father was a construction worker who was having trouble holding a job.

"She couldn't talk and she was functioning like a child of just a year when we found her," said Weisz. "She's a beautiful child but she's having a hard time relating to anyone any more."

The silent little girl is just part of a trend that is growing here and in other energy boom towns springing up in the West from Montana to New Mexico. As the frenzied rush to cash in on the huge coal deposits takes hold, experts like Weisz are becoming alarmed; the great energy rush is already taking its toll among boom-town children. . . .

Part of the problem, according to several experts, is that parents in boom towns are often young themselves—the average age here is under 25—and have never learned to accept the responsibilities of rearing children.

. . . one young father in Rock Springs returned to his trailer recently after 16 straight hours on a construction job and found his 6-month-old infant crying. "The man decided the best way to shut the kid up was to feed it vodka. . . . The baby drowned in its own vomit."

Source: Bill Richards, "Trailer Houses, Row on Row—In Most You'll Find Woe on Woe." *The Salt Lake Tribune*, December 19, 1976: B4.

Federal Regulatory Agencies

The initial impetus for conducting social impact assessments came in response to expanding regulatory demands associated, at least in part, with the upsurge in environmentalism that occurred in the late 1960s and early 1970s. In 1969 the U.S. Congress passed the National Environmental Policy Act (NEPA), which requires comprehensive, systematic evaluation of the effects on the natural and human environment of major federal actions or actions taken by other agencies that affect federal lands or other entities over which the federal government has regulatory or management responsibility.

In effect, this act formalized a role for social scientists to assess the social (in addition to environmental and physical) impacts of projects. Rules

ILLUSTRATION 12.2 An example of trying to remedy an unanticipated impact.



"Well, if you can't use it, do you know anyone who *can* use 3,000 tons of sludge every day?"

© Sidney Harris. From *What's So Funny about Science?* Los Altos, Calif.: Kaufmann, Inc. Reprinted with permission.

and regulations were established by the Council on Environmental Quality (CEQ) to implement the policy outlined in NEPA. The regulations developed by CEQ explicitly note the need to consider *social* as well as ecological and other impacts of a project. The regulations and policies developed during the following decade ensure that the social effects of major federal actions or actions utilizing federal resources are identified, and that amelioration of the negative effects are considered in the decision-making process (Branch et al., 1982).

State and Local Government Concerns

Following the passage of NEPA in 1969, many state and local governments became concerned about the assessment of impacts of projects in areas under their jurisdictions. Most of the direct impacts of a large-scale project are highly localized. As a consequence, the state and local agencies that are responsible for providing services for the incoming population or that will face regulating and monitoring a project have a stake in the accurate measurement and projection of impacts. The state and local

interests are understandable when one considers the potential costs of providing housing, police protection, medical care, public schools, and so on for a large immigrating work force in a small community with extremely limited revenue and tax base. Imagine a community of 400 or 500 residents providing the necessary services for 10,000 construction workers and their families who will stay in the area for only four or five years. As a result, state and local concerns about accurate impact assessment have often been tied to the need for strategies to finance the new facilities and services required in a community because of the proposed development. One goal, of course, is to identify potential sources of support from industry and the federal government to help meet the costs associated with the new development, costs that simply cannot be borne by the impacted community.

Because of such concerns, a number of states have adopted environmental and/or facility siting legislation that imposes impact assessment requirements similar to those imposed by NEPA (Auger and Zeller, 1979; Leistriz and Murdock, 1981). In fact, several states have developed impact assessment requirements that go beyond NEPA, particularly in requiring extensive impact monitoring for the duration of a project and the development of a detailed mitigation plan before the project is allowed to begin. The mitigation plan usually describes what industry must do in providing funds to develop the required services and facilities.

Industry Concerns

Just as state and local governments are increasingly interested in social impact assessment, so is industry. There are several important reasons for this. First, because of the state and federal laws noted above, impact assessments are required to obtain the necessary permits for a proposed development to proceed. When a social impact assessment is poorly done, it is open to challenge by federal and state agencies and by project opponents. Such actions have led to long, costly delays in approval for projects.

Second, unmanaged and unplanned-for boom growth has often led to community problems that have adversely affected the ability of industry to attract and retain a stable work force (Metz, 1980). As a result, private industry has a clear stake in assessing potential impacts accurately and working with communities to alleviate the most serious problems.

Third, industry has become involved in social impact assessment to facilitate siting by relieving the anxieties of residents of the affected communities. The negative reports about the problems of boom towns have contributed to resistance in some communities to the siting of a large energy development in their vicinity. The desire to reduce such concerns has led industry to try to build greater community support for proposed developments.

Finally, some states require industry to accept financial responsibility for the adverse impacts on a community associated with their project (Leistritz and Murdock, 1981; Watson, 1977). In such cases, industry is interested in determining exactly what impacts they will be responsible for and what the costs of amelioration are likely to be.

For all of these reasons, there has been a dramatic increase in interest in social impact assessment. It has become an essential part of the planning, permitting, impact monitoring and mitigation processes associated with most large-scale industrial development.

III. WHAT IS INCLUDED IN A SOCIAL IMPACT ASSESSMENT

Social impact assessment is a part of the larger **enviromental impact assessment** process. The information gathered and analyzed by the social scientist becomes part of a larger report that includes data about possible impacts on the local ecosystem, including the economy, animal and plant life, air and water quality, agricultural production, and so on. Social impact analysis is also closely related to and often integrated with the economic impact analysis. When this is the case, attention is typically directed toward determining five types of impacts, which are summarized in Illustration 12.3 and include the following:

1. *Economic impacts*—including changes in business activity, jobs and employment, personal income, and so on.
2. *Demographic impacts*—including changes in regional, county, and community population and population characteristics such as age structure, sex composition, and so on.
3. *Fiscal impacts*—including the level and distribution by jurisdiction (county, city, school district) of public costs and revenues such as providing water treatment plants, courthouses, jails, etc. and taxes and bonds to pay for them.
4. *Community service impacts*—including changes in demand, distribution, and quality of such community services as schools, health care services, water and sewer services, police and fire protection, transportation, and social services.
5. *Social impacts*—including changes in community organization, community perceptions, lifestyles and satisfaction, and the effects of the proposed development on such specific groups as the elderly, minorities, and people living on fixed incomes. (Murdock, Thomas and Albrecht, 1982)

Strictly economic impact assessment (items 1 and 3) are not considered in this chapter. Instead, we focus on demographic, community service, and social impacts, including all of them under the more general heading "social," as distinct from "economic." Nevertheless, it should be recognized that there is direct interaction between social and economic impacts.

ILLUSTRATION 12.3 *POTENTIAL COMMUNITY IMPACTS ASSOCIATED WITH A LARGE-SCALE DEVELOPMENT*

<i>Economic/Business</i>	<i>Community Services</i>
Increased local employment	Demand for more services
Higher wages	Demand for better service
Increased business sales	quality and accessibility
Higher cost of living	Demand for more and specially
Competition for labor	trained personnel
Development of new business	Improvement of service facilities
Increased competition for	and equipment
resources	Higher cost to residents for
Possible economic losses after	services
the project is completed	Improvement of area road and
Increased land values caused	rail systems
by growth	Increased demand for housing
Changes in uses of land	
<i>Demographic</i>	<i>Social</i>
Increased population due to new	Perceived changes in quality of
workers and families	life
Changes in the location of popula-	Increased social problems
tion growth	Value conflict between new and
Changes in the age, income, and	longtime residents
educational characteristics of	More formalized interaction among
the population	residents
	Additional sources of community
	leadership
	More active public participation
	in predevelopment of site
<i>Fiscal/Government</i>	
Additional sales tax revenues	
Increased governmental payments	
to communities	
Increased property taxes	
Increased cost of community	
services	
Reassessment of property values	

Source: Adapted from S. H. Murdock and F. L. Leistritz (1979).

Factors Affecting the Nature and Magnitude of Social Impacts

To assess the social impacts associated with a given project, the researcher usually focuses on two sets of factors: site area characteristics and project characteristics. *Site area characteristics* determine such things as the nature and type of resources that the community can draw upon to deal

with change. The importance of community resources seems obvious. The availability of space in the schools, adequate housing, and good medical care are critical in estimating probable impacts (Branch et al., 1982). Also important is the community's experience with previous projects. Other important site area characteristics include composition of the community's population; its geographical location and distance from other communities that can provide workers, services, and facilities; labor force; recreational facilities; other services; fiscal resources; political organization; and attitudes of community residents about change (see Illustration 12.4).

The second major set of factors that determine impacts are the *project characteristics*. A project brings people into a community along with jobs, income, resources, new organizations, and regulations. A sudden change in population may alter local health and safety conditions if, for example, local crime rates go up. Six characteristics of a project need to be determined:

1. The number and demographic characteristics of the people who will enter or leave the community.
2. The number and types of jobs created and their general distribution among longtime residents and newcomers.
3. The amount and general distribution of income brought into the community.
4. The magnitude and type of resources brought into the community (including tax revenues).
5. The number and type of organizations and regulatory systems brought to bear upon the community.
6. Changes in health and public safety. (Branch et al., 1982: 3-4)

If one is able to accumulate the necessary background information on both the impact area and the project, then the task of determining the nature and types of potential impacts is fairly straightforward. We will now describe how that social impact analysis is actually done.

IV. DOING SOCIAL IMPACT ASSESSMENTS

A standard set of social impact methods has been relatively slow to develop. In fact, techniques of social impact assessment are still evolving. In the remaining sections of this chapter, we will describe the current "state of the art."

It is recommended that, rather than a single method for social impact assessment, different approaches be used. The recent handbooks recommend the use of at least three different approaches (see Branch et al., 1982; Murdock, Thomas and Albrecht, 1982): secondary data analysis, survey research, and participant observation.

ILLUSTRATION 12.4 Major factors determining impacts of a project.**A.** Size Area Characteristics

Natural resource base
 Current pattern of resource utilization
 Population characteristics
 Size
 Age structure
 Employment characteristics
 Training and skill levels
 Unemployment and underemployment
 Social and cultural homogeneity
 Geographical location and isolation
 History of development
 Availability of housing
 Public services and infrastructure
 Schools
 Police and fire protection
 Health and medical care
 Social services
 Recreation
 Water and sewer
 Local attitudes, values, and preferences

B. Project Characteristics

Work force requirements
 Site
 Skill requirements
 Project timing and duration
 Construction phase
 Operations phase
 Direct and indirect population impacts
 Characteristics of the incoming population
 Age structure
 Marital status
 Similarity with existing population in such factors as
 ethnicity, religion, etc.
 Amount and distribution of new income
 Project resources

The ideal is not possible in all cases. In some situations, one may be forced to rely more on a single method or upon limited aspects of one or two methods. Murdock, Thomas, and Albrecht (1982) identify the critical factors in selecting methodological approaches as the scope of information required, the availability of existing data on these topics, and the resources required by an ideal combination of approaches in contrast to the resources available to the researcher.

The usual starting point is to determine exactly what information is

needed. Different projects require different types of information, but in virtually all cases information is needed about the characteristics of the local population, the local public service delivery system, and the community context.

Data availability must be considered. Information about the size and characteristics of the local population can be obtained from U.S. Census Reports and government planning agencies. If the relevant secondary data describing the impact area are not available, the researcher will have to depend on primary data-collection techniques.

Finally, the decision about an assessment technique is greatly affected by the availability of resources, including money, time, and personnel. A detailed community ethnography may take several years to complete, while a social impact assessment must often be completed in six months or less. Similarly, interviews with a random sample of community residents may be very costly and the researcher may not have the budget to complete such an analysis. All of these factors are weighed in selecting a research methodology.

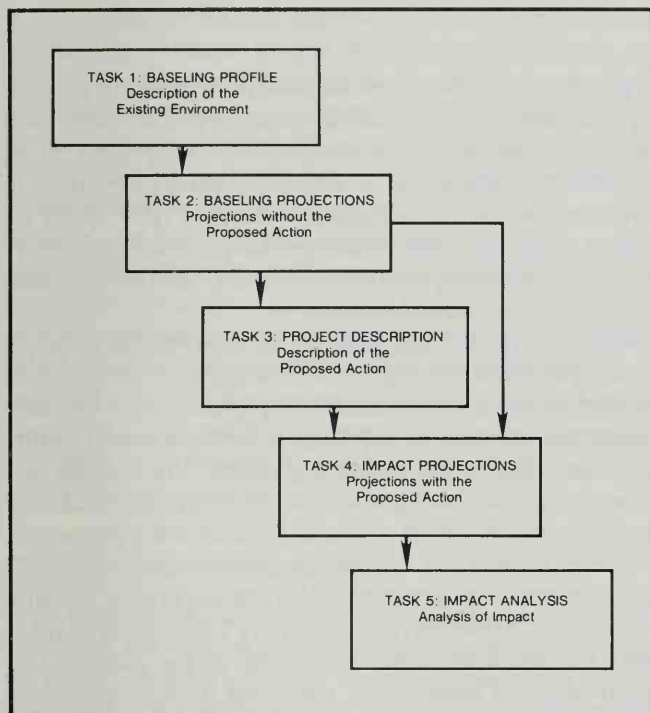
Defining the Task

In its simplest form, social impact assessment involves five distinct tasks. These are outlined in Illustration 12.5.

Baseline profile First, the researcher must complete a comprehensive **baseline profile** of the community or region to be impacted. This profile, a comprehensive description of the study area, includes how many people live in the area and a general description of their age, occupational skills, education, employment characteristics, relevant ethnic and religious characteristics, and so on. It also includes a detailed description of the public service sector of the community, number of schools, number of students, number of law enforcement officers, jail facilities, number of doctors, bed capacity of hospitals, recreation facilities, water and sewer system capabilities, and so on. Attention is also given to the local social organization: the degree of integration and whether there are any political, ethnic, and religious divisions in the community. The attitudes of local residents toward growth and change are also assessed.

Baseline projections The second major task is the development of a series of projections of the degree and direction of change in each of the factors described in the baseline profile without the proposed project. The task is to predict how the area is likely to change *without* the project. For example, if a community has experienced the outmigration of most of its young people because there are few employment opportunities, the baseline projections might reflect a continuation of this trend. An analysis of

ILLUSTRATION 12.5 Steps in the impact assessment process.



Adapted from Branch et al., 1982.

the age structure of the local population will permit forecasts of school enrollments and the demand for new facilities and personnel.

Project description Task three in the assessment process is to provide a detailed description of the proposed action. For example, the description will detail how many workers will be required and when they will arrive in the community. The length of the construction and how the worker demands differ in the construction phase from the operating phase will be documented.

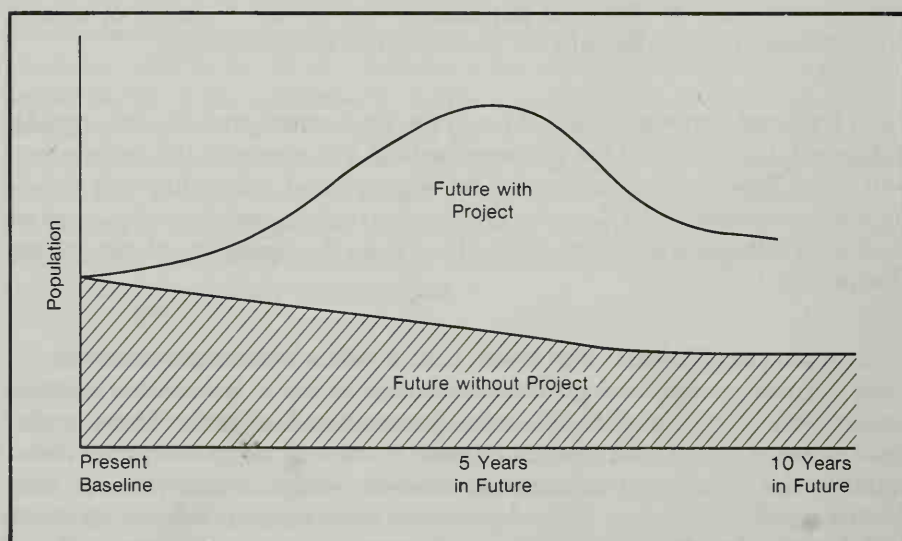
Impact projections The fourth task involves a detailed discussion of how the project will change the community or area. It includes making responsible estimates of the following parameters: how many of the workers are likely to be hired locally and how many will migrate into the community; the number of immigrating workers and how many wives and children will accompany them; how many new houses, schools, doctors and hospital beds, police officers, and water and sewer systems will be

required. Potential conflicts between oldtime residents and newcomers are also anticipated.

Impact analysis Finally, the differences between the baseline projections and the impact projections are assessed to indicate the amount and kinds of changes likely to derive from the proposed development. It is important to recognize that the researcher is not comparing current conditions with a set of future conditions. Important changes may occur in the community and region independently of the special project being studied. Thus, the researcher is comparing two sets of future conditions—one with and the other without the proposed project.

Illustration 12.6 outlines hypothetical changes in a community's population with and without a potential development. The baseline projections indicate a continuing decline in the community's population. The *with project* figures show a dramatic increase in population during a construction phase, followed by a decline during the operating phase. Predictions are made as to whether the social well-being and quality of life of residents would remain unchanged, improve, or get worse as a result of the proposed project. The most probable answer, of course, is "none of the above." Quality of life is likely to improve in some ways but to decline or remain stable in others. For example, better-paying jobs may become available, contributing to a higher standard of living. However, at the same time, local crime rates may go up, and these same residents may be afraid to

ILLUSTRATION 12.6 Hypothetical changes in community population with baseline and impact projections.



walk the streets at night. Moreover, the positive and negative impacts of a proposed action are not likely to be equally distributed across a given population. Trained young people may benefit economically from the project; the same may be true for the local businesses. However, older residents on fixed incomes may not reap any of the financial gains and, at the same time, may see their purchasing power decline as a result of inflation and taxes and may see their community change in other ways they do not like. Adjustments occur easily in some areas. Given an expanded tax base, new schools can be built to accommodate the new students. However, social system adjustments may not occur so easily and the maladjustment may contribute to serious social problems. Cortese and Jones (1977) note that people commit suicide because of inadequate social systems, not inadequate sewer systems. Moreover, the adjustments in some sectors may increase the maladjustments in others. In other words, for some residents, the formal, tax-supported mechanisms developed to replace the former informal mechanisms do not work nearly as well as did the "outmoded" systems they replaced.

A Triangulated Approach

Conducting a social impact assessment is not simple and we recommend the use of a triangulated approach. Several techniques may be necessary to deal with the variety of questions that face the social impact researchers. A comprehensive social impact assessment will probably require secondary data analysis, survey research, and participant observation.

Using secondary data The first task in almost any impact analysis is to describe in detail existing conditions in the community or impact area. This description provides the researcher with the baseline against which to measure changes that occur as a result of the proposed action. Much descriptive information about a community is available in secondary sources and can simply be copied or transcribed.

Secondary data analysis depends on data collected previously for other purposes. Therefore, the researcher must remember that the data were not generated and compiled with impact assessment in mind. Nevertheless, such data are typically among the key ingredients in a social impact assessment.

Although secondary data sources are used most frequently to *describe* the characteristics of an impact area and to identify important historical trends, they may also be used for more specific impact analysis. For example, secondary data can be used to identify the types of social impacts most likely to occur and to specify the meanings they may have for the community. If in an analysis of secondary data on local schools, one determines that there is no excess capacity in existing school facilities, that

there are no plans approved for new facilities, and if the proposed action will result in the addition of new families with school-aged children, then it follows that an important impact will be overcrowded schools.

The baseline profile is typically based on data obtained from secondary sources. These serve as the *basis* for projecting the type and magnitude of impacts that will occur in the community. That is, the baseline information about the community is combined with project data to determine the nature and magnitude of the impacts to be attributed to the project.

The experiences of previously affected communities, described in secondary sources such as published impact statements, can help identify and clarify the relationships between project inputs and community change. Information from secondary sources may also be used to check data obtained in other ways.

The problems with using secondary data are described in Chapters 9 and 10. Briefly, they include the fact that the data were collected for another purpose; the interests of the data collection agency may differ from those of the social impact researcher; much of the information the researcher needs may not be available and some of the information that is available may not be presented in the desired form; and data on residents' attitudes, values, perceptions, and interaction patterns are usually not available in secondary data sources. In addition, much secondary data are badly dated and may not reflect recent changes in the community. Thus, while secondary data analysis is a critical part of the impact assessment process, it must almost always be supplemented by other data collection efforts.

Illustration 12.7 summarizes the major sources of secondary data that can be used in social impact analysis and describes some of the types of information available in these sources.

The decennial census of population of the United States is one of the richest sources of information about a community. The following types of information are available from the census at the county level: county population size (changes in population can be determined by examining previous censuses), the age and sex distribution of the population, educational levels, ethnic makeup of the population, data on migration (movement into and out of the county), occupational distribution, employment by industry, family characteristics, and income distribution.

In addition to the census of population, a decennial census of housing is conducted that provides information on such things as housing type, facilities, ownership patterns, and dollar values of rented and owned housing. Every five years (years ending in 2 and 7—i.e., 1982, 1987, etc.), the Bureau of the Census publishes the City and County Data Book, which summarizes economic, demographic, and governmental services for all counties and cities with populations greater than 25,000. Censuses of agriculture, mining and manufacturing, and businesses are also conducted regularly and may provide some useful baseline data.

ILLUSTRATION 12.7 Sources of secondary data for community impact analysis.

DATA SOURCE	TYPES OF INFORMATION AVAILABLE
U.S. Census	Extensive demographic data including age, sex distribution, education, ethnicity, migration patterns, etc.
Bureau of Labor Statistics	Extensive information on such things as employment, unemployment, types of employment, income, etc.
National Center for Health Statistics	Information on vital rates such as births, deaths, health, etc.
Federal Bureau of Investigation	Data on crime and crime rates by category
State Department of Health	Birth and death data, morbidity rates, numbers of hospitals and hospital beds, numbers of doctors and nurses, number of nursing homes and beds, marriage and divorce rates, etc.
State Employment Commissions	Number employed by industry, projected levels of employment growth, available jobs skills and skill shortages
State Industrial Development Commissions	Labor availability by city, type and location of industrial firms, housing availability, etc.
State Highway Departments	Miles of highways and streets, condition of highways and streets, capital and maintenance costs for highways
State Law Enforcement Agency (State Police or Department of Public Safety)	Number and type of motor vehicles, types of crimes and violations, number of police officers by county and city, law enforcement costs by jurisdiction
State Outdoor Recreation Agency	Number and type of parks, number and type of campgrounds, location and use rates for parks, lakes, rivers, etc.
State Welfare or Human Services Department	Number of families on various types of assistance such as Aid to Families with Dependent Children, Social Security, and SSI Size and staff qualifications of state and regional agency offices including number of social workers, number of alcohol and drug abuse counselors, number of family counselors Number and cases of child abuse, spouse abuse, desertions, child adoptions

ILLUSTRATION 12.7 *continued*

DATA SOURCE

TYPES OF INFORMATION AVAILABLE

State Fire Marshal's Office	Number of fires, location of fire departments, fire personnel by type (volunteer or professional) and location, fire equipment by location
Secretary of State's Office	Political precincts and jurisdictions, elected officials by area, voting patterns
Environmental Impact Statements	Frequently, environmental statements will have been prepared on other projects in the area. These can provide valuable data on many of the variables noted above.
Local Histories	Local authors frequently have written local histories that can be helpful in developing a broader understanding of the area and its people.
Local Newspapers	Scanning local newspapers is an excellent means to become better acquainted with a community and its principal actors as well as the issues that have been of greatest local concern.

Two problems with census data are of special concern to the social impact researcher. First, most of the readily available census publications provide data only at the county level. If the site of primary impact is a single community or several communities in a large county, the countywide data are sometimes not very useful. Second, census data are dated even when first published. If one is doing the study near the end of a decade, available census reports may not describe important recent changes, a crucial problem for communities characterized by rapid social change.

In addition to the census of population, several other federal agencies and publications provide information that can be used. For example, extensive information on employment and unemployment is available from the Bureau of Labor Statistics, and the National Center for Health Statistics provides information on vital rates such as births, deaths, and health. The Federal Bureau of Investigation publishes summaries of reports from local law enforcement agencies. Often the accuracy of data from such sources can be checked with community and county planners, local law enforcement officials, civic leaders, and educators. State governments also publish annual vital statistical reports that contain statistics on marriages, divorces, populations, births, deaths, crime, health, and infant mortality. Such reports can usually be obtained by visiting appropriate agencies in the state capitals.

In addition to offering state reports, many states are divided for ad-

ministrative and other purposes into multicounty Associations of Governments. Frequently, these multicounty organizations have collected extensive data for the counties over which they have responsibility. They will often have the most current and accurate information on population, employment and housing, and sometimes detailed information on such things as facility and service capacity and availability that have been accumulated for planning and grant application purposes.

Other important sources of secondary data for the social impact researcher are other impact statements written previously in the area. Many new projects are planned for areas where previous projects already exist. For example, a coal-burning electrical power generating plant may be planned for an area where there has been large-scale strip-mining. Researchers may be able to obtain extensive information about the impacted community from the previous studies. Contact with local planning agencies and with federal agencies such as the Bureau of Reclamation, the Bureau of Land Management, the National Park Service, and the Forest Service can usually direct the researcher to impact statements that may have been prepared previously.

Finally, to get a better picture of the community, its background, and the issues and concerns that are most important to its citizens, the social impact researcher can usually benefit from a visit to the local library and local newspaper offices. A review of local histories and a scanning of the newspaper files is an excellent way to become acquainted with the community and can help the researcher understand local issues and identify the key actors in the community who must be contacted for other phases of the research process.

One important caution must be expressed about this use of secondary data in social impact assessment. One does not collect information from a secondary data source simply because it is available. Sometimes inexperienced researchers will present an array of data in the impact assessment *not* because these data best answer the critical questions but because they were available. Consequently, some social impact statements contain much information that is superfluous or redundant and that does not apply directly to the task at hand. The researcher must be careful always to keep the key research questions in mind and to obtain the information necessary to answer these questions.

Using survey methods Previous chapters in this book have described sampling and interview and questionnaire surveys. Many of the questions the social impact researcher must answer simply are not treated in available data. Consequently, some primary data collection is almost always required. Data about attitudes, values, and preferences of local residents generally are collected via survey. Also important information about the interaction and lifestyle patterns of community residents that helps the

researchers understand the likely impacts is most effectively obtained via some form of survey research. Murdock, Thomas, and Albrecht cite several advantages of this methodology for the social impact researcher. These include:

1. A well-designed community survey will produce information that can be generalized or applied to the community at large.
2. The survey method can be used to quantify a wide variety of attitudinal, perceptual, and behavioral information by providing questions structured in ways to facilitate the classification of responses. This capability is crucial because social impacts involve many dimensions of community life, and the project impacts must be measured and integrated with other quantified socioeconomic data.
3. Compared to many other methods, a survey permits greater speed. In small communities which are likely to be unprepared and to lack the resources to deal with the rapid changes accompanying a large-scale development, early information may be vital to planning.
4. A survey will produce objective data relatively free from the perceptual biases of assessment specialists. (Murdock, Thomas, and Albrecht, 1982: 34-35)

Survey methods also have certain disadvantages that may limit their usefulness in the social impact process. Among the most important of these are the following:

1. A survey obtains information from community respondents at a single point in time. Different types of impacts occur in a community at different times during a project. They will vary in intensity, scope, and duration—all of which connote changes in communities over time. Subsequently, attitudes may change drastically as impacts from a project occur and intensify.
2. The survey may be inappropriate for addressing some types of issues, such as very controversial ones. The utility of the survey is especially problematic when interviewers have only a few moments to win the confidence of each respondent and to motivate him/her to participate in the study. Interviewers must be highly skilled to conduct surveys and maximize participation and at the same time to probe for attitudes about controversial issues.
3. Communities may become over-surveyed. Residents may become annoyed and reluctant to give candid responses.
4. A survey is limited to the assessment of the responses to questions that are asked. As a result, surveys may fail to address some key issues. (Murdock, Thomas, and Albrecht, 1982: 36-37)

Perhaps the most critical problem associated with the use of survey research in social impact assessment, however, is the fact that the Office of Management and Budget (OMB) has placed some rather severe restrictions on surveys that are conducted with federal support. Specifically, **OMB Circular A-40** states that one cannot use a questionnaire that asks the same questions to more than nine persons without first obtaining OMB clearance on the instrument. Obtaining such clearance has often been an extremely

laborious and time-consuming process and therefore has not been feasible within the time constraints. Researchers have usually gotten around the letter of the law by using highly unstructured interviews based on "topical guides" which avoid the problem of their asking the identical question to more than a few individuals. The guide will usually contain a series of general issues that the researcher wants to cover in the interview but the manner and sequence in which the questions are asked will vary from interview to interview. A second strategy to deal with this restriction is to interview key or knowledgeable informants. This may entail conducting interviews with a relatively small number of community leaders who are assumed to be knowledgeable about various points of view in the community.

Four basic questions must be answered by the researcher who uses survey methods in social impact assessment (Branch et al., 1982): (1) What information is needed? (2) Who has that information? (3) How can it best be obtained? (4) How should the data be analyzed to provide the findings needed by the decision-maker?

The first question involves determining the questions that the social impact researcher must answer that have not been answered previously. The survey will usually aim to obtain data from community residents on attitudes, perceptions, and related issues that both affect and will be affected by the proposed project.

The second question refers to identifying the potentially impacted group. Although it may be desirable to interview a sample of community residents, that might not be feasible given time, budget, and OMB constraints. Therefore, the researcher may be limited to obtaining survey data from a few community leaders or residents. If it is important to know something about the attitudes and preferences of all community residents, and if one has the budget and if OMB restrictions do not apply, then it will be important to use some form of probability sampling, as described in Chapter 3. If the researcher is not interested in the precise estimation of the characteristics of the population and if the data will be used to inform the investigation rather than to test specific hypotheses (Branch et al., 1982), then nonprobability sampling procedures may be used.

If one is limited by resources or other constraints, then it may be necessary to forego any attempt to obtain data from a sample of the general population. In such cases, the researcher can use a form of key informant interviewing. Information might be obtained from "leaders" or "experts" who are assumed to be able to speak for the larger community. Respondents might be formal leaders of the community (selected because of the positions they occupy) or informal leaders (selected because of their reputation for knowledgeability or local influence). In either case, it is important that the researcher obtain information from several points of view. For example, if only formal community leaders are interviewed, then the

position of minority groups, the elderly, or individuals who might oppose a particular project will often not be represented.

It is sometimes useful to use a **snowballing** technique to identify key informants. With this procedure, one begins with a list of names obtained by using either a **positional** or a **reputational** (who really knows "what's going on") **approach**. Each person interviewed is asked to identify others who would know a lot about the issue or who should be contacted because they speak for important segments of the community. Used effectively, this snowball procedure usually brings the researcher into contact with most of the informal and formal leaders of a community and allows him or her to identify all of the important issues and concerns of different groups of community residents. It is important to recognize, however, that one cannot generalize from the information obtained in interviews with key informants to residents of the entire community.

The third question to be answered by the researcher—how the information is to be obtained—involves four basic techniques: face-to-face interviews, telephone interviews, mailed questionnaires, and questionnaires that are dropped off and picked up by the researcher. As was discussed earlier, face-to-face interviews are a very effective but costly way to obtain large amounts of data. The interview can be based on a highly structured interview guide or it can be rather unstructured, guided by only the researcher's desire to obtain certain types of information. When unstructured interviews are obtained, they are usually conducted in a conversational manner and responses might not even be recorded until after the interview. Usually, however, the researcher will take notes during the interview or tape-record it. Face-to-face interviews allow the researcher to probe, to explain, to follow up on important points that are raised by the interviewee, and to obtain detailed and rich information. The primary disadvantage is cost, which usually limits the number of interviews that can be afforded. The researcher will want to at least conduct a few interviews with informants in the community to supplement the data obtained in other ways.

Because of costs of face-to-face interviewing, many have turned to telephone interviews. Such interviews have two major advantages: lower costs and speed. The elimination of travel time can dramatically cut the costs of data collection and can increase the number of interviews that can be conducted within a specified period of time. The primary disadvantage of telephone interviews in social impact assessment is that they are by nature brief. Therefore, many of the issues that the researcher may wish to explore cannot be treated. One commonly used procedure is to conduct relatively brief telephone interviews on a few of the most critical issues with a sample of community respondents and then to conduct more detailed face-to-face interviews with a sample of key informants.

Mailed questionnaires require the creation of highly standardized re-

search instruments and are feasible only when OMB approval is not required or has been obtained. However, a mail questionnaire survey obtains large amounts of data from large samples of respondents at relatively low cost. Since acceptable response rates will often require three or more follow-ups to the initial mailing, the researcher must plan carefully to ensure that the data needed are available within the allotted time.

Another strategy is for the researcher to deliver the questionnaire and then return at a later time to retrieve the completed instrument. This procedure is more costly and time-consuming than a mailed survey but allows personal contact and gives the researcher the opportunity to explain the study, answer questions, and solicit participation in the study.

Regarding the fourth basic question, the type of analysis used in an impact assessment is dictated largely by the data collection method(s). If quantified data have been obtained from samples of community residents, then the analysis will include coding the responses, creating data files for machine analysis, "cleaning" the data, and so on. An impact statement is a type of *applied* social research report, and the results are usually descriptive of characteristics illustrated in simple tables.

However, if the data derive from unstructured, informal interviews, then little statistical analysis is possible, and the researcher will use the illustrative qualitative data from the "expert" informants to supplement the statistical data obtained by other means. For example, crime statistics for the community can be supplemented by impressions and firsthand experiences of police officers, obtained in interviews. Both qualitative and quantitative data are necessary to complete a readable, responsible social impact statement.

Using participant observation A third procedure frequently used along with secondary data analysis and survey research in social impact assessment is participant observation. As with the other data collection methods, the emphasis that this procedure receives varies from project to project. However, on all projects the researcher will need to spend time in the community collecting secondary data and interviewing local leaders and some time can often be spent in participant observation.

As was discussed in Chapters 4 and 8, participant observation is one of the oldest research methods used in the social sciences. The researcher deliberately involves himself or herself in the community being studied in an attempt to gain an insider's perspective on events, gathering data by participating in the daily life of a group or organization (Becker and Geer, 1957: 28).

With most social impact analyses, the available time and resources limit the amount of participant observation that can be done. Rarely is there time for a formal or even a brief but systematic ethnography. Lacking the resources for a thorough study, the impact assessor participating in aspects

of community life has the primary goal of experiencing some "flavor" of life, activities, attitudes and values of the community's residents. In addition, observation can be combined with the interviews, questionnaires, secondary statistics, and expert opinions to provide greater triangulation. The selective reporting of critical events, meetings, conversations, and so on by the well-trained observer carries an aura of factuality—a sense of how things really are at the grass roots level that is often missing in the dry summaries of statistics and "expert" opinions.

In the simplest sense, what the social impact researcher does in the community is to become involved in the everyday life of that community for a brief time in order to better understand the community from the perspective of its inhabitants. A major advantage is the very fact that he or she is in the community to understand the potential social impacts of a specific proposed project. Consequently, we are not talking about the typical community ethnography where the researcher must study a very broad array of issues and concerns. The general problem is already defined, and while new ideas and problems will emerge in the field, the researcher cannot allow these to intrude upon the specific task at hand.

Before going into the field and during the early stages of the field visit, the researcher should become familiar with the community and also learn something of its history and of the interests and concerns of its people. This acquisition of background can usually be accomplished by visiting the local newspaper and library, by conducting some informal interviews with local officials and leaders, and by viewing the community via extended walks or drives so that landmarks, neighborhoods, major roads, and community institutions become familiar. The participant observer's role in the impact assessment context will then basically involve visiting, observing, and talking to people in their natural setting. Usually the role of the researcher as an employee of a government agency, university, or private consulting firm is sufficient to provide access to local "experts" and community leaders. There is no need to proceed under false identities. At least in the United States, the social impact researcher has become a familiar participant in the planning and negotiations among local, county, state, federal, and private interests.

There are relatively few situations where it is desirable or even possible for the social impact researcher to be a "complete participant," if only because his or her connection with the impact analysis project should be known to most persons in the community. In addition, most impact researchers who serve as participant observers for a time are quite aware of their continuing research role and of their own minimal long-term stake in the community life they must observe.

As is true with other research methods, often the first step in doing participant observation is to obtain some form of permission or at least toleration from elite persons or "gatekeepers" in the community. Failure

to obtain cooperation from these influential people may result in the researcher's not being accepted by community residents generally. In the process of obtaining information from these persons, the researcher can often gain tacit or even explicit approval and support from them.

Whether community leaders are interviewed or not, it is often helpful for the researchers to spend time among community residents without conducting interviews or using other structured data-gathering procedures. The purpose of this unstructured visibility is to allow people in their natural context to feel comfortable with the researcher. Another reason for the unstructured participation is that, when interviewing, the researcher almost always originates the action and the local resident reacts. In unstructured participation, the researcher observes subjects as they initiate their own action independent of stimuli provided by the researcher.

It is suggested that one start with an effort to understand the organization of the community and its history, then turn to understanding day-to-day routines and interactions, and finally, focus on people's more private feelings, attitudes, and sentiments. The difficult part of the participation process is deciding what is worth recording or what merits further investigation, and the knack for being right in such decisions is as much an art as a science.

The Importance of a Triangulated Approach

In this chapter we have stressed the use of a triangulated approach in doing social impact assessments. The three methods of secondary data analysis, survey research, and participant observation have been discussed in terms of their specific applicability to the impact assessment process. The relative advantages of the three approaches are summarized in Illustration 12.8.

It should be noted that the strengths of one method often offset the weaknesses of another. Also data that are available via one of these methods may not be available through any other approach. The combination of the three approaches minimizes the disadvantages inherent in any one method and maximizes the likelihood that the final product will reflect the diversity inherent in human communities while underscoring basic themes or common perceptions of consequences.

The task of integrating the three approaches can be rather straightforward (Murdock, Thomas, and Albrecht, 1982). For example, one can begin by collecting relevant secondary data and using these data sources to develop an initial community baseline profile. The baseline profile will include information on factors such as population and employment characteristics, educational level, age structure, and so on. This information provides guidance as to what other types of data are essential, either by identifying gaps in existing knowledge or by highlighting topics that seem

ILLUSTRATION 12.8 Advantages and disadvantages of the three methods used to assess the social impacts of a proposed development.

METHOD	ADVANTAGES	DISADVANTAGES
Secondary Data Analysis	-Provides information unavailable from other sources (e.g., resource base and ethnic mix). -Data collection costs are relatively low. -Respondent participation requirements are minimal. -Less highly skilled research staff can be used for data collection.	-Incapable of providing some essential data (e.g., perceptions and levels of interactions). -Key sources of variation in individual perceptions cannot be determined. -Data analysis requires analytic sophistication and access to computer facilities. -Secondary data may be dated and not reflect recent changes. -Reliance on secondary data collected for other purposes means it may not be relevant for social impact assessment.
Survey Research	-With appropriate sampling, information can be generalized to the entire community. -Attitudinal, perceptual, and behavioral information can be quantified and statistically analyzed. -Permits rapid data collection and administrative ease in collecting and reporting data. -Surveys can be easily replicated. Comparable data can be obtained from several communities. -Bias can be minimized.	-Data are gathered at a specific point in time. Change is difficult to monitor. -Getting cooperation may be difficult, especially if a survey deals with sensitive issues. -Communities near project sites may become over-surveyed. -Obtaining OMB approval may be difficult and time-consuming. -Some issues of concern to community residents may not be addressed by the survey.
Participant Observation	-Allows for an in-depth understanding of events or issues. -Allows the researcher to observe community residents in their natural setting.	-Potential for loss of detachment and loss of objectivity. -Researcher is directly exposed to the subjects, which may be emotionally difficult.

ILLUSTRATION 12.8 *continued*

METHOD	ADVANTAGES	DISADVANTAGES
	-Method is flexible and allows the researcher to be surprised.	-Potential for floundering where the researcher obtains little information of value.
	-Is not affected by OMB restrictions on questionnaires and interviews.	-Data collection may suffer from problems of reliability and validity.

Source: Adapted from Murdock, Thomas, and Albrecht (1982: 147–148).

to merit more intensive research. For example, questions that should be asked in the community surveys can be identified, and background information can be accumulated that will assist the researcher in participant observation.

The next step is usually to make initial community contacts and schedule a field visit. The field visit is used to collect data by all three research methods. Contacts are made with local officials and agency heads to obtain additional secondary data (for example, contact with the school superintendent can be used to obtain the latest figures on school enrollments, facilities, and capacities), initial key informant interviews are conducted, and a general entree to the community and its organizations can be obtained for the purpose of subsequent participant observation.

Next, the researcher usually constructs questionnaires and interview guides and makes decisions about sampling for the survey work, prepares a list of contacts for key informant interviewing, and continues to become involved in the community life in order to observe firsthand residents' reactions to and attitudes about the proposed action.

Since one or a few researchers often must "wear all of the different research hats," coordination among the different data-collection efforts is not as difficult as in many other types of research involving multiple methods. Gaps in data obtained by one method can be filled by another (for example, by conducting an additional interview or spending some time with representatives of another group in the community). The final result should be a detailed, complete data base adequate for preparing an accurate, timely social impact statement that is useful to those who are to decide whether, and how, a proposed project should proceed.

V. IMPACT MITIGATION AND MONITORING

In recent years the role of the social impact researcher has expanded beyond that of completing a social impact statement to be used in decision making. It is now recognized that to deal effectively with the social impacts that

occur, it is necessary to continue to monitor the project and to develop strategies that will help mitigate or change potential negative consequences of the project.

As we noted earlier, several states have now passed industrial siting legislation as an outgrowth of the initial NEPA legislation at the national level. This legislation often requires that an **impact mitigation program** be worked out and approved before permits are granted for development to proceed. Mitigation measures include a broad array of actions designed to minimize a proposed project's negative effects. Such actions may include varying construction schedules to avoid "peaks" and "valleys" in population impacts, adopting local hiring policies in order to minimize the number of workers that will have to migrate to the area, maximizing employment opportunities for local people, and providing monies to the community. These monies are provided in advance of the project's completion and attainment of profit-making capacity so that the community can expand the existing facilities and services to meet the needs of the population, by means of, say, the construction of a new school or hiring additional law enforcement personnel or social service workers. Upfront monies may be provided by industry through the prepayment of taxes or through grants made to the community.

Previous experience in impacted communities has shown that many of the adverse effects of large-scale developments can be reduced by effective mitigation measures (Branch et al., 1982). Therefore, the social impact researcher must be able to collect and interpret information necessary to make recommendations for mitigation programs.

The primary task of the social impact researcher in the creation of effective mitigation programs is to identify, during the course of the research project, things that might be changed to minimize the negative effects of the project and then to determine how such changes might be made. The impacts of a project can usually be altered in two ways—by changing the characteristics of the project, or by changing the characteristics of the affected community (Branch et al., 1982). Potential changes that might be identified during the social impact research might include the need for financial input from the company to help build new housing to avoid a serious housing shortage, the development of a program to welcome newcomers to the community so that a division of residents into "newcomers" and "old-timers" does not occur, or the provision of government or private funding of additional mental health workers or recreation specialists.

Impact monitoring is done as a follow-up to the social impact assessment. It involves the task of continuing to collect data while a project is being implemented in order to ensure that the projected changes do occur, to assess whether the expected benefits from the project actually seem to be occurring, and to see that unexpected adverse effects do not

occur, or, if they do, that they are immediately identified and dealt with. If present trends continue, it will become a mandated, institutionalized process for all major projects. Because the monitoring effort is closely tied to the initial impact assessment effort, monitoring community change is likely to become an important specialty in social research.

VI. SUMMARY

Social Impact Assessment does not involve the creation of a new method but is concerned with the application of a variety of techniques to a particular type of problem: that of assessing the impacts of a proposed action on a community or region and its inhabitants. A diversity of projects now require formal social impact assessments. The construction of a new nuclear power plant, the opening of a strip-mining operation, the building of a hydroelectric dam, or the construction of a new water storage and transmission system are common examples. All such projects potentially have important positive and negative effects on the people living in proximity to the proposed development and on those who will migrate in for temporary work or to remain as permanent residents. A social impact assessment is an attempt to estimate the nature and scale of the positive and negative impacts that may be anticipated.

Social impact assessments are now required by federal mandate for any federal action that affects the natural and human environment or for any other action that affects federal land or other entities over which the federal government has regulatory or management responsibility. Following the passage of federal legislation requiring social impact assessments, many state governments enacted their own legislation requiring some form of impact assessment prior to the granting of permits for projects to begin construction or development. Industry has also become involved in the assessment process in an effort to help mitigate potential negative impacts that a project might have and at the same time to identify in advance probable costs to the industry of dealing with those impacts in legal ways that do not alienate influential segments of the local or national public. For all of these reasons, the role of the social scientist in social impact assessment has increased significantly in recent years.

Social impact assessments usually include efforts to identify the demographic, the community service, and the quality of life impacts of a proposed action. The nature and magnitude of these impacts can usually be estimated by two sets of factors: the characteristics of the proposed project and the characteristics of the impacted community.

The process of social impact assessment involves five related tasks: (1) developing a comprehensive baseline profile of the community or region that is to be impacted; (2) creating baseline projections, or determining

how the area would likely change *without* the proposed project; (3) producing a detailed description of the proposed action; (4) making impact projections, or determining how the proposed project is likely to change the community or region; and (5) assessing the differences between the baseline projections and the impact projections. The differences between these two sets of figures (the baseline or "without" projections and the impact or "with" projections) are used to estimate the amount and kinds of changes likely to be associated with the proposed development.

The most effective way to do social impact assessment is usually some form of multidimensional or triangulated approach. We proposed using a combination of secondary data analysis, survey research, and participant observation. Each of these methods has been described in greater detail in earlier chapters. In this chapter their special application to the process of social impact assessment was described.

There is a growing opportunity for social researchers in the design of creative modes of mitigating the negative effects of projects, and in monitoring change as development occurs. Impact *mitigation* involves not only identifying actions that can be taken to decrease the negative impacts of a proposed project, but also finding ways to increase the positive impacts. Often during the course of a social impact assessment the researcher will be able to identify actions that would benefit both the project and the community, and alternative courses of action and their probable consequences should be specified in the social impact statement. Impact *monitoring* involves collecting data during the life of a project to determine if the projected impacts occur as anticipated. The monitoring effort is closely tied to the initial impact assessment process, in part because the initial research provides the baseline against which change is measured.

REFERENCES

- AUGER, C. S., and MARTIN E. ZELLER, 1979. *Siting Major Energy Facilities: A Process in Transition*. Boulder, Col.: The Tosco Foundation.
- BECKER, HOWARD S., and BLANCHE GEER, 1957. "Participant Observation and Interviewing: A Comparison," *Human Organization* 16: 28-32.
- BRANCH, KRISTI, JAMES THOMPSON, JAMES CREIGHTON, and DOUGLAS HOOPER, 1982. *Guide to Social Assessment*. Prepared for the Bureau of Land Management Social Effects Project by Mountain West Research, Inc., Billings, Mont.
- CORTESE, CHARLES F., and BERNIE JONES, 1977. "The Sociological Analysis of Boom Towns," *Western Sociological Review* 8: 76-90.
- DEUTSCHER, IRWIN, 1973. *What We Say/What We Do*. Glenview, Ill.: Scott, Foresman.
- FREUDENBURG, WILLIAM R., 1979. "People in the Impact Zone: The Human and Social Consequences of Energy Boomtown Growth in Four Western

- Colorado Communities." Unpublished Ph.D. dissertation, New Haven, Conn.: Department of Sociology, Yale University.
- . 1982. "The Impacts of Rapid Growth on the Social and Personal Well-Being of Local Community Residents." Pp. 137-170 in Bruce A. Weber and Robert E. Howell (eds.), *Coping with Rapid Growth in Rural Communities*. Boulder, Col.: Westview.
- GILMORE, JOHN S., and MARY K. DUFF, 1975. "Socioeconomic Impact Management: Are Impact Assessments Good Enough to Help?" Pp. 278-298 in J. M. McDonald (ed.), *Computer Models and Forecasting Socio-Economic Impacts of Growth and Development*. Edmonton, Alberta: The Faculty of Extension, University of Alberta.
- LANTZ, ALMA E., and ROBERT L. McKEOWN, 1979. "Social/Psychological Problems of Women and Their Families Associated with Rapid Growth." Pp. 42-54 in U.S. Commission on Civil Rights (ed.), *Energy Resource Development: Implications for Women and Minorities in the Intermountain West*. Washington, D.C.: U.S. Government Printing Office.
- LEISTRITZ, F. LARRY, and STEVEN H. MURDOCK, 1981. *The Socioeconomic Impact of Resource Development: Methods for Assessment*. Boulder, Col.: Westview.
- LOVEJOY, STEPHEN B., 1977. *Local Perceptions of Energy Development: The Case of the Kaiparowitz Plateau*. Los Angeles: Lake Power Research Project, Bulletin Number 62.
- McKEOWN, ROBERT L., and ALMA E. LANTZ, 1977. *Rapid Growth and the Impact on Quality of Life in Rural Communities: A Case Study*. Glenwood Springs, Col.: Colorado West Regional Mental Health Center.
- METZ, WILLIAM C., 1980. "The Mitigation of Socio-Economic Impacts by Electric Utilities." *Public Utilities Fortnightly* (September 11, 1980): 3-11.
- MOEN, ELIZABETH, ELISE BOULDING, JANE LILLYDAHL, and RISA PALM, 1981. *Women and the Social Costs of Economic Development: Ten Colorado Case Studies*. Boulder, Col.: Westview.
- MURDOCK, STEVEN H. and F. LARRY LEISTRITZ, 1979. *Energy Development in the Western United States: Impact on Rural Areas*. New York: Praeger.
- MURDOCK, STEVEN H., JOHN K. THOMAS, and DON E. ALBRECHT, 1982. *Handbook for Assessing the Social and Special Effects of Nuclear Waste Repository Siting*. A Report to the Office of Nuclear Waste Isolation, Battelle-Columbus and the Department of Energy.
- PATTINSON, MICK, ROBERT WEISZ, and ELIZABETH HICKMAN, 1979. *Program Evaluation Reports: A Study of the Impacted Population, Related Stresses, Coping Styles, and Mental Health Program Needs in the Gillette, Wyoming Planning District*. Sheridan, Wy.: Northern Wyoming Medical Center.
- WATSON, K. S., 1977. "Measuring and Mitigating Socioeconomic Environmental Impacts of Constructing Energy Projects: An Emerging Regulatory Issue." *Natural Resources Lawyer* 10: 393-403.
- WEISZ, ROBERT, 1979. "Stress and Mental Health in a Boom Town." Pp. 31-47 in Judith A. Davenport and Joseph Davenport, III (eds.), *Boom Towns and Human Services*. Laramie: Department of Social Work, University of Wyoming.

Analyzing Data

thirteen

I. Introduction

II. Coding

III. Data Entry

Cleaning the data

IV. Descriptive Analysis

Levels of measurement in descriptive analysis

Frequency distributions and comparisons

V. Cross-Tabulation and Contingency Tables

VI. Testing Relationships Between Variables

Tests of significance

nominal and ordinal data

interval data or data with different levels of measurement

Measures of association

nominal data

ordinal data

interval data

VII. Indexing and Scaling

Thurstone scales

Likert scales

Guttman scales

VIII. Summary

I. INTRODUCTION

Up to this point we have focused on the methods used to collect data relevant to some particular research problem. We have seen that the social scientist has available a variety of data-collection tools. The selection of certain strategies from among these tools is influenced by a variety of considerations, including the nature of the research problem, the resources available, the purpose for the research (for example, to test causal theory or to do exploratory work on a new problem), the extent to which data already collected are available or whether primary data must be collected, and so on. The outcome of the application of any of the methods, however, is the accumulation of research data relevant to the research problem. The next critical task is the analysis of those data to make sense of what has been collected and to answer the questions that prompted the research. Several previous chapters have made reference to the analysis phase of research. Here we will discuss the steps involved.

Data analysis is what one does with the questionnaires, interviews, documents, experimental data, field notes, or other data collected during a research project. It is the stage of a project in which one tries to answer the questions, "What have we found? What do the data reveal?" Analysis usually follows the completion of data collection and precedes (and in some measure overlaps) the writing and reporting of results. In the outline of

stages in research (Chapter 2) the process of data analysis was said to include procedures to convert the data into standard forms that would simplify computations and comparisons, and the application of analytical techniques that would produce results that met the objectives of the project.

Data analysis may itself be divided into five stages and in this chapter we consider each in detail. The stages are: First, **coding**, in which verbal responses or written answers are converted to numbers to facilitate their handling. Second comes **data entry**, in which the coded data are punched on computer cards or, as is more often the case nowadays, are entered directly onto computer tapes or disks. The data file thus created is "cleaned": the entered data are checked item by item (that is, variable by variable) to be sure that there are no illegitimate or impossible responses; the key-puncher or computer operator verifies the entry by entering the data twice, and the two data files are compared, thus minimizing the probability that entry errors were made that are not obvious impossible or illegitimate punches. The third stage is **descriptive analysis**, which usually refers to looking at how responses to individual items are distributed or, in more technical terms, to examining the frequency distributions for individual variables. The fourth is **cross-tabulation**, in which the relationships between two or more variables are examined. Fifth is **testing relationships between variables**, including the reduction of multiple indicators of concepts to manageable summary measures through **indexing and scaling**. An essential process not considered in detail here is interpretation, a process that occurs during both the analysis and report-writing stages of a project. Interpretation consists of assigning meaning to the findings and deciding what conclusions are justified.

II. CODING

Coding is the method whereby responses to items in a data-collection instrument—a questionnaire, an interview guide, a schedule for content analysis, or an observation sheet from an experiment—are converted to standard form, generally but not necessarily a numerical form, so that systematic analysis may be done by data-processing machines or electronic computers. The data-collection instrument may contain many modes of response, ranging from check marks for one of a series of numerical responses, through brief verbal descriptions of one's occupation, to qualitative material such as a respondent's statement about personal attitudes or behaviors, to detailed observations which vary in length from a few words to several paragraphs. The coding process—a literal translation of the respondent's answers to a standard numerical language—is made possible by a **codebook** that assigns a range of numbers for each variable corresponding to the various response categories. The code-

book also contains detailed instructions about how to handle exceptional responses so that different coders using the same codebook and the same completed instrument should, ideally at least, produce identical translations.

Some researchers construct their questionnaires, code sheets, or observation forms in such a way that a minimum of coding is necessary. "Precoded" response categories enable a circled or checked number on the answer sheet to be entered directly into a data file. Thus the data instrument itself serves many of the functions of a codebook. However, even precoded instruments require a supplemental codebook for at least two reasons. First, instruments are designed, or at least *ought* to be designed, to make the respondent's, coder's, or observer's job as easy as possible. Instrument format, numbering of items, and the like that make it easy for a layperson to answer or the researcher to record observations often make the direct transfer of responses from the instrument to a computer card or magnetic tape more difficult. Second, there is usually a need for instructions to coders on what should be done with responses that do not fit in the format provided in the printed instrument, or with inconsistent or inappropriate responses.

The ideal codebook provides for each variable or item of information in the data-collection instrument an unambiguous guide for translating any conceivable response into a systematic, numerical system. Normally a codebook includes a complete reproduction of the instrument. The items are worded exactly as they were on the instrument, but there is additional information telling the coder (or keypuncher, if the data are being entered directly) what numbers are to be entered in ambiguous or unusual cases as well as providing a clear guide for the typical responses.

Let us illustrate some of the techniques of codebook construction and coding by looking at portions of one of the codebooks from a questionnaire survey conducted during the Middletown III project. The data consisted of questionnaires administered to students in Middletown's high schools by their home room teachers. Portions of the first four pages of the codebook for this high school survey are reproduced in Illustration 13.1.

The upper-left corner of the page identifies "CARD 1" as the location of the data described below. This label appeared on each page of the codebook until 80 columns (or possible digits) had been accounted for. Then the label is changed to "CARD 2" for the next 80 columns, and so on. These data were originally punched on 80-column IBM cards before being transferred to magnetic tape, and because there were more variables than could be contained on one card, each page of the codebook identified which of the seven cards necessary to encode all of the data was being described.

ILLUSTRATION 13.1
Sections of the Middletown III High School Survey Codebook

CARD 1			
Column	Variable	Question	Code
6-7	1	1	AGE How old are you? _____ (If 1/2 used, code to next 99. No response <u>year; ex: 16 1/2 is 17</u>)
			SCHOOL CLASS What is your class in school? (please circle the number in front of the appropriate response on this and similar questions) 1. Freshman 2. Sophomore 3. Junior 4. Senior 9. No response
8	2	2	
...	...	...	...
Do you belong to any of the following organizations or extracurricular activities, either in the school or in the community? (Be sure to circle a response for every item "a" through "o"). If all left blank, code as 9's; if one or more 2's are circled and all rest blank, code the blanks as 1's.			
			ATHLETICS Athletic teams 1. No 2. Yes 9. No response
12	5	5a	
			BAND OR ORCH Band or orchestra 1. No 2. Yes 9. No response
13	6	5b	
...	...	...	...
			SCHOOL TRACK What high school curriculum programs are you presently following? 1. General 2. Business 3. College prep 4. Vocational 5. I don't know 6. If college prep & another <u>7. If business & vocational</u>

ILLUSTRATION 13.1
continued

CARD 1			
Column	Variable	Question	Code
31	22	10	8. Other combinations 9. No response
...	...	...	...
GRADES What kind of grades did you receive on your last grade card? 1. A's or mostly A's 2. A's and B's 3. Mostly B's 4. B's and C's 5. Mostly C's 6. C's and D's 7. Mostly D's or below 9. No response			
47	38	12	
If more than one response is circled, code the response that is the closest to the average of all circled responses. If too mixed, code as 9, "no response."			
...	...	...	
CONFIDANTE TYPE Code the "Yes" responses to the preceding question, adding them to the codebook as you go. 00. Answered "No" to the preceding question. 99. Answered "Yes" but did not specify the type of confidante. 01. Girlfriend(s) 02. Boyfriend(s) 03. Best friend(s) 04. Friend(s) 05. Brother			
...	...	...	
17. Cousin 18. Mother, Friend(s) 19. Girlfriend, Minister			
...	...	...	
157. Mother, Cousin 158. Girlfriend, Brother, Minister, Aunt, Sister 159. Girlfriend, Boyfriend, Mother, Sister			

In the illustration observe the capitalized headings, COLUMN, VARIABLE, QUESTION, and CODE. Under COLUMN is entered the number of columns (or possible digits) devoted to the variable described. Under VARIABLE is a unique identification number for each variable in the data set. The codebook for the high school survey describes 298 variables. The heading QUESTION refers to the identifying number or letter of the question producing the data as it appeared on the questionnaire. Some questions produce data for more than one variable, and some questionnaires renumber within sections, and therefore there is no necessary correspondence between the variable number and the question number. The question number is retained in the codebook to simplify reference to the questionnaire. Some codebooks do not include this information, but we have generally found it convenient to have the questionnaire reproduced exactly in the codebook, with item numbers as well as the item itself given in full.

Finally, the word CODE heads a column in which is found a brief label for the variable in question, followed by the question and response categories, if any, as given on the questionnaire. There are also additional coding categories as needed to account for all responses and nonresponses. Sometimes there are special instructions and coding categories that emerged as necessary during the coding. We have underlined these special instructions and categories to distinguish them from the categories and instructions that appeared in the original draft of the codebook before it was amended to include the additional data.

Continuing with Illustration 13.1, note that the first four columns are not variables but are set aside for the identification number of the questionnaire. In the Middletown high school survey over 1,600 completed questionnaires were coded and analyzed, and each questionnaire had a unique identifying number stamped on it, thereby assuring that the responses of each individual would consistently be linked to him or her and no other. The first case was identified as 0001, the second as 0002, and so on. All ID numbers fit within the initial four-column field. The fifth column was left blank.

The first variable described in the codebook, entered in columns 6 and 7, was the student's answer to the question, "How old are you?" Only two columns were allowed for this variable, and we were only interested in the students' age as generally reported, that is, age at last birthday. A few students provided more detail than requested, giving answers like "16½," and therefore a special instruction, the underlined segment of the variable description, was added to ensure that all coders would treat such responses the same way.

The possible confusion in coding this initial item leads us to a basic principle of coding, namely, *decisions about problematic responses must be consistent*. The issue is not so much whether a student aged 16½ should

be coded as 16 or 17, but rather that, whatever decision is made, it should be applied consistently.

A final point to be noted in the coding instructions for variable 1 in the illustration is the "no response" option. Of course this category was not printed on the questionnaire; to have done so would have greatly increased the number of students who did not answer questions. But even on straightforward items such as age or gender, there are respondents who leave the items blank or answer inappropriately so that their responses are, in effect, nonresponses. Accordingly, for each variable a "no response" option was designated from the outset. Any number may be used to denote nonresponse—some codebooks use a zero or leave the column blank—but a consistent code for nonresponse simplifies coding. In this Middletown survey, the number 9 was used to indicate nonresponse, except for variables where 9 or 99 or 999 might be meaningful responses. In those few instances some other numbers, perhaps zeroes, were assigned to nonresponse.

Now refer to the second variable, "School class." This characteristic proved to be straightforward enough that, except for the nonresponse option, no additions or special instructions were necessary. The item and four response categories appear in the codebook precisely as they did in the questionnaire.

The third entry in Illustration 13.1 is variable 5, "Athletics." "Athletic teams" were among 15 types of extracurricular activities or organizations for which there was a specific query about membership. In the questionnaire there was a general question, which applied to the following 15 items. In the codebook the general question was extended across the entire page instead of being indented in the manner of questions that generated a single answer. Each type of organization was then listed as a question generating a distinct variable, even though the probe printed in the questionnaire might be only a single word, such as "Dramatics" or "Debate." This entry in the illustration is interesting for another reason. It shows how, despite the detailed instruction, "Be sure to circle a response for every item 'a' through 'o,'" there were respondents who circled 2's for organizations they belonged to but left the other possibilities blank instead of circling 1's as instructed. The researchers had to decide whether an activity/organization that was left blank represented a "no answer" (9) or a "don't belong" (1) response. It was decided that if *all* of the 15 activities were left blank, the respondent had probably not answered the question—at least he or she had not followed the instructions at all—and therefore variables 5 through 19 were coded 9 for no response. However, if for one or more of the activities the respondent had circled a 2, denoting membership, and no 1's had been circled, we assumed that the respondent was simply being "lazy," marking the organizations he or she belonged to—

following the instructions halfway—and leaving unmarked the organizations to which he or she did not belong. The operating principle that produced this decision was: *never throw away data; if there is evidence that a respondent has responded consistently and intelligibly, although not in accordance with instructions, coders may "correct" responses to conform to the standard format.*

Another coding principle is apparent in the special instructions added to the variable "School track." The questionnaire offered five possible responses, "General," "Business," "College prep," "Vocational," and "I don't know." These proved to be insufficient for a few students, who marked two or more of the five options. When coders encountered multiple responses to this item, we decided to add categories capturing what seemed most essential about the responses, namely (1) did a combination include the high-status alternative, "College prep?" Ideally, each combination encountered would have generated a new code. However, only one column had been allocated for this item, and adding new codes for the combinations including "College prep" (6) and the "business and vocational" plan (7) left only one additional possibility (8). Accordingly, that was assigned as a residual code for "Other combinations." There were so few students in this residual category (.7 percent of students answering the question) and the inconvenience of redefining "School track" as a two-column variable long after most of the questionnaires had been coded, dictated the decision to use 8 as a residual category. Besides, the basic theoretical distinction we were likely to be concerned about in using the variable was the distinction between college-bound students and others. That distinction had been captured in the designation of College prep and any other response(s) as code 6. Fine distinctions beyond that did not seem to justify the costs of reformatting the variable.

The principle at issue is: *Coding decisions are pragmatic as well as theoretical, and sometimes a category system that captures most of the variation in a set of responses is preferable to a more detailed but costlier system that, in the opinion of those directing the research, is not likely to matter much.*

The necessity for code-building as the coding proceeds is apparent in the last question in Illustration 13.1, "Do you have someone you can confide in and tell your troubles to, someone who understands you?" The precoded response categories were "No" and "Yes." Students who answered yes were asked to "please explain who, *not by name* but by type of person, such as girlfriend, uncle, brother, minister, etc." The codes for the responses given had to be exhaustive enough to account for every respondent. At the same time, they had to allow for the possibility that some respondents would have many confidants of different types.

Note that the codes given make provision not only for persons who had answered "no" to the previous question (coded 00) but also for those who had answered "yes" but did not answer the probe about type of confidant (coded 99). We created a unique two-digit code for each different

type or combination of types reported. When in the course of coding it turned out that there were over 100 different types and combinations, we added a residual category (coded 98) for all codes not covered in the first 99 possibilities and added later in the data file a detailed three-digit code for all respondents given a 98. Note that in this case, the fine detail was deemed important enough that information "lost" in the residual category was added elsewhere so that the combination of the two data fields (columns 54-55 on Card 1 and columns 64-66, formerly blank on Card 4) captured all of the details about confidants offered by the respondents. Having each combination coded separately allows researchers to combine categories to fit whatever analytical categories are appropriate. For example, in one phase of the analysis, we divided the respondents into two classes: all those who had mentioned one or both parents as a confidant and the remainder, who had not mentioned at least one parent.

Beyond the initial obvious categories for types of possible confidants—girlfriend, boyfriend, brother, sister, etc.—specific codes were generated by the responses given. Thus the first questionnaire that contained the combination "girlfriend and minister" as close confidants was assigned a separate number (19), and all codebooks were updated to include that. In a questionnaire coded much later, a respondent again mentioned girlfriend and minister, but also listed three other types, "brother, aunt, and sister." This new combination of five types of confidant was assigned a new code (158).

If during the analysis we wished to know how many students listed a girlfriend as a confidant, all codes containing the term girlfriend, including numbers 19 and 158, were combined in a single category. Similarly, if we were interested in separating students who had a minister as a confidant from others who did not, again punches 19 and 158, plus any others that included minister alone or in combination with other types of confidants, were grouped in a single category.

Two basic code-building principles illustrated in the coding of type of confidant are (1) *to be exhaustive, a code need not include all logical possibilities, but it must include an unambiguous code for all possibilities that in fact appear in the data set*, and (2) *where possible, retain maximum detail in assigning code definitions*. It is much easier to assign very detailed codes and combine them during the analysis than to have to return to an original data set for a new coding operation because the available details were overlooked or ignored in original coding. One never knows what subsequent analytical problem will be more readily manageable because fine detail in coding categories captured all of the response variation provided by the subjects of the study. Coders must necessarily participate in code-building when they encounter a response not assignable to any of the categories available to them and, after appropriate consultation with their supervisors, add a new category. However, it is essential that coding instructions be complete enough that

coders do not have to puzzle about which category a response fits, or to make impressionistic "shoot from the hip" decisions. A good codebook contains sufficient detail that the coder's responsibility is limited to assigning a response to one of several clearly defined categories. Coding should be mechanical in that it merely consists of sorting responses among categories that are defined so well that no strenuous decision making is required. If coders have to agonize over individual responses, then the codebook is inadequately prepared. Ideally, the process of coding is comparable to that of a conveyer belt carrying units of information that pass through a mechanical sorter programmed to recognize the essential differences in the units and quickly allocate each into one of several available bins. Continuing the analogy, ambiguity of definition or inadequate detail in coding instructions results in the sorting machine "tilting" or "burning out" because it is not programmed to face the kinds of discriminations it suddenly confronts. The ideal is never reached—coding never becomes entirely a matter of mechanical sorting according to instructions that handle all exigencies—but a good codebook reduces most coding to unambiguous sorting rather than complicated, nonreplicable decision making.

A principle that applies both to instrument construction and coding is that respondents should not be expected to be accurate coders. Questions should be framed to produce sufficient information that trained coders can make appropriate decisions; one never knows what biases may affect a respondent's decisions about which category his or her own special characteristics fit best. For example, it is usually a mistake to ask people about their occupations by presenting them with a series of broad occupational categories and expecting them to check the one that fits their job. Many questionnaires include an item on occupation that reads something like this:

"Which of the following best describes your present occupation? (Check one)":

- ☐ 1. Professional or technical worker
- ☐ 2. Manager or administrator
- ☐ 3. Sales worker
- ☐ 4. Clerical worker
- ☐ 5. Craftsman or foreman
- ☐ 6. Operative
- ☐ 7. Laborer, except farm
- ☐ 8. Farmer or farm manager
- ☐ 9. Farm laborer or farm foreman
- ☐ 10. Service worker, except private household
- ☐ 11. Private household worker
- ☐ 12. Other (please explain) _____

This set of categories corresponds to the gross occupational categories used by the U.S. Census Bureau. The problem is that the questionnaire or

interview schedule never contains sufficient detail for either an interviewer or a respondent to know where many specific occupations belong. A man whose title at the factory where he works is "technician" may wish to classify himself as a technical worker rather than an operative, and the distinction between operatives and laborers is often unclear. The category "service worker" is also problematic; deciding whether what one does is a service, a craft, or a technical skill is too complicated to leave in the hands of amateur coders.

A far better system for obtaining occupational data is simply to ask the respondent his or her occupational title, a brief description of duties that go with the job, and the industry the job is in. Trained coders can then take this information and, using a guide such as the Census Bureau's (1971) *Alphabetical Index of Industries and Occupations*, determine one's occupational category. This index contains the titles of over 23,000 occupations and the designation of which specific occupational category each occupation has been assigned to by the professionals at the Census Bureau. The result of this process is an easily manageable three-digit identification number for any occupation, and the coding process does not depend on a coder's familiarity with a wide range of occupations. The *Alphabetical Index* has been constructed with reference to the occupational diversity of the entire nation, and greatly reduces coding mistakes. The coding principle exemplified in this rather lengthy example is that *respondents should never be expected to be coders*. The more complicated the eventual set of categories, the more individual, discrete detail should be sought in order that professional coders will have sufficient information to code an entire data set consistently.

Recent developments in the **software** available for computer analysis have made it convenient to enter not only numbers but also word-for-word statements or even lengthy quotations from instruments that combine both quantitative and qualitative items. Entry of such data may be done in two ways: if a paragraph describes a respondent's attitude toward religion, a coder might decide that the paragraph reflected one of a handful of major themes, each being identified by a single number. Another technique, more acceptable because it requires less decision making by coders, is to enter the entire statement, verbatim. Later, all qualitative comments for a given category of persons may be scanned by the skilled analyst especially trained in extracting common themes. The latter technique is also useful in finding illustrative comments to flesh out or add human interest to the numerical results of an investigation. For example, in the Middletown survey mothers with children were asked how they would feel if they came to believe that there was no God. Entry of their word-for-word responses would have permitted sorting by race, age, marital experience, religious denomination, frequency of church attendance, or a combination of these. Then for each specific type of respondent the qualitative responses given to the question

could be printed out along with the identification number of that respondent, so that patterns in or contrasts between the responses of mothers in the various categories could be highlighted through systematic comparison of the uncoded, word-for-word responses to the open-ended question.

This technique is also very useful when one wishes to identify the demographic characteristics associated with the respondent making a quotable statement. It is frequently useful to the analyst and meaningful to the eventual "consumer" of research to know that a particular quotation is from a person identified by age, sex, marital status, occupation, or whatever demographic variables are appropriate.

III. DATA ENTRY

Data entry involves transferring the code representing a response on a questionnaire from field notes, etc. to a computer card, tape, or disk. As we have said, coding and data entry may be performed in a single operation, provided that the coding instructions are simple enough that the person entering the data can make decisions quickly while transferring the data to the IBM card or magnetic tape. In our experience, any coding that requires consulting a source other than the codebook, or that requires time-consuming cross-referencing in the codebook itself, is best done as a separate operation, with the codes written directly on the questionnaires. The keypuncher can then simply enter the items that required special coding at the same pace as the simpler precoded items.

An illustration of this procedure is the coding of occupations, often the most difficult part of a coding operation. Coding occupations in a consistent, correct way usually requires reference to an intermediate source like the Census Bureau's *Alphabetical Index of Occupations and Industries*, described earlier. We have obtained the best results by training one or more "specialists" in occupational and other complicated coding and having these experts do all of the coding necessary for those complex items before passing the instruments on to the keypuncher.

Cleaning the Data

After a codebook is prepared, the data set coded, and the codebook revised along the way as appropriate, the data are punched onto cards or entered directly onto magnetic tape or disks. The researcher is then ready for the final stage of data preparation, cleaning and verifying the completed data file. Cleaning a data set refers to searching for coding errors that are identifiable by being impossible or improbable given the way the variables are defined in the codebook. Each variable has an acceptable range of

responses. For example, a set of detailed codes for religious preference might range from 01 to 54, meaning there were at least 54 specific denominational categories that appeared in the data. There might be a 55th category for persons who said they "didn't know," and 99 might denote all "no answer" questionnaires, where a respondent had left the religious preference item blank. Thus, the acceptable range of responses on this item would include 01 through 55, and 99. Any entry between 56 and 98 is an illegitimate response, as would be a blank and a zero.

There are now computer programs that will list the identification number of every case that shows an entry falling outside the legitimate boundaries of the variable. With this list of "errors," the analyst then looks at the questionnaire, determines what kind of error produced the punch, and then corrects the data file.

The difficulty with such cleaning programs is that they can identify only those responses that are out-of-range. Mistakes that are within range—for example, a case where a Catholic was coded as a Protestant—do not show up as errors because both punches are legitimate. The usual way to minimize such coding errors is to enter the data twice and match the resulting cards or tapes. This process is called "verifying the data." If computer cards are being used, the keypuncher punches a card and then repunches the same card. He or she is immediately alerted if the second punching differs from the first. In direct entry onto magnetic tape or disk, the simplest though somewhat expensive method is to enter the entire data set twice and then have the computer match the two data files item for item and list all instances where there is a discrepancy between the two files. Then the researcher goes back to the original questionnaires and determines what the correct number is. Such verification by double-entry also minimizes errors that might result from a variable's being skipped, which would result in large blocks of the data being incorrect.

Researchers may do a series of reliability checks on the people who enter their data and, finding that a given keypuncher makes so few errors that the duplicate punching is unnecessary, may decide to avoid the additional cost of double entry. However, unless one has documentation that a keypuncher is sufficiently accurate, based on a large number of previous entry jobs, the double entry is an essential part of cleaning the data.

Modern computer systems facilitate direct data entry by having the entire questionnaire entered so each item appears on the screen of the console in sequence. The data-entry person enters the answer appearing for that item, and the next question then appears on the screen. Because the entry system is programmed not to accept an out-of-field response, it is unnecessary to clean this type of error. However, there is no guarantee that within-field errors were not made, and so verification through double entry is still necessary.

IV. DESCRIPTIVE ANALYSIS

Having entered and cleaned the data, the analyst is ready to examine the characteristics of the sample and to begin looking for relationships. The next stage of analysis is the production of frequency distributions or “**marginals**” (the totals that appear in the margins of a tabulation) which describe the distribution of answers to each item or variable in the data set. A **frequency distribution** is simply a list of all the possible categories for each variable, showing the number of respondents in each category. Normally frequency distributions as presented in computer output show both the numerical and percentage distributions for each variable. Inspection of these “univariate” or single-variable distributions is essential both for deciding what the next steps in the analysis ought to be and for establishing cutting points or category limits when the original detailed entries are too numerous for meaningful additional tabulations.

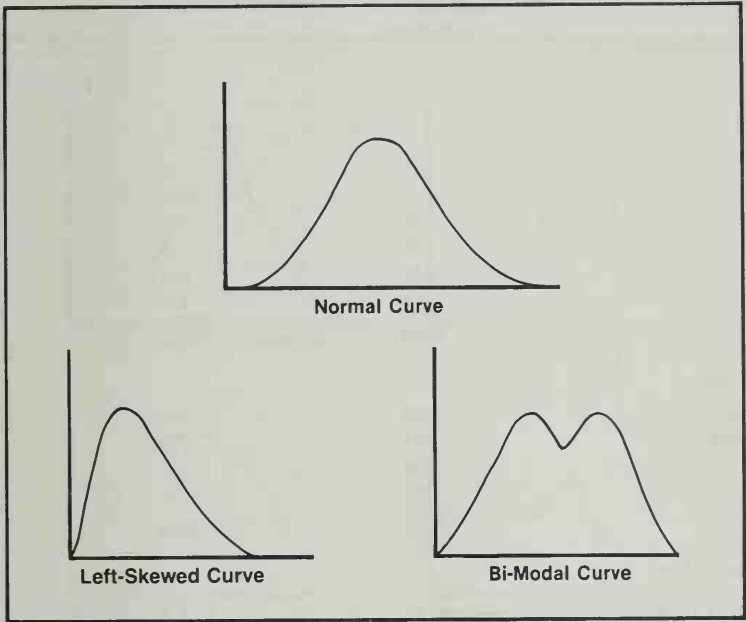
Inspection of the frequency distributions one variable at a time permits the analyst to decide whether a characteristic is rare or frequent, and whether it is distributed “normally”—with most people falling in the middle and very few at either extreme—or whether it is distributed in some biased or skewed fashion, such that very few people have low values and almost everyone has high values (or the reverse) on the variable in question. Examples of normal, skewed, and multimodal univariate distributions are shown in Illustration 13.2.

The analysis of how scores on any variable are distributed usually includes a description of the range over which the scores are distributed and identification of the most frequent score (sometimes referred to as the **mode**), the **median** score (the score that represents the middle of the distribution, with the same number of cases having higher values as have lower values), and the **mean** or arithmetic average, calculated by summing all of the scores and dividing by the number of cases for which there are scores.

Consider first the variable, “Age,” shown in the first panel of Illustration 13.3. The range of ages represented among the Middletown High School students runs from a precocious 11-year-old through 19-year-olds. The modal, or most common age, is 15. The average, or mean age, is 15.9, and the median is also 15.9 (methods for computing the median from grouped data are discussed in elementary statistical texts and will not be described here). These measures of central tendency—mode, mean, and median—which characterize any dispersion of scores are generally printed along with the numerical and percentage frequency distributions in the output generated by most computer programs used in the preliminary analysis of social science data.

The second panel of the illustration shows the distribution for the variable “School Class.” Here the range is from freshman to senior, and

ILLUSTRATION 13.2 Normal, skewed, and bimodal curves.



the modal response is freshman. If we define the categories in terms of number of years of schooling, with freshman equaling 1 and senior equaling 4, then the mean is 2.3 and the median is 2.2. In other words, the “average” student in the high school sample was a sophomore, i.e., had completed 2.3 years of schooling, although the single most common class (the modal category) was freshman.

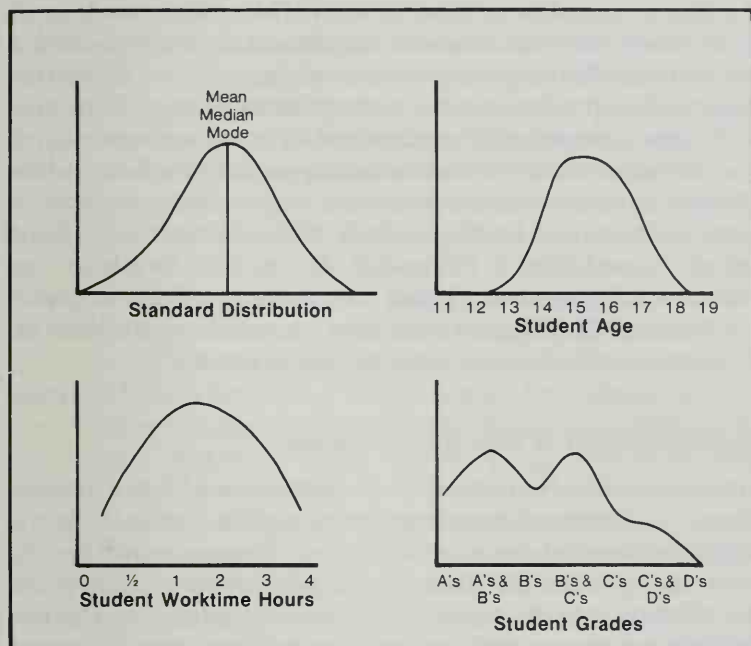
The study of the shape of the frequency distributions for age and school class is facilitated by charting the *percentage distributions*. Just as there are ideal or standard measures of length (the meter or the yard) and weight (the gram or the pound), so there are standard types of frequency distributions against which an obtained distribution may be compared. The one used most often is called the **standard normal distribution**, and it is reproduced in the top panel of Illustration 13.4.

Note that a key characteristic of the standard normal distribution is that the mean, median, and mode are all at the same point, at the very peak of the distribution, and that the percentage of cases that fall to the left of that peak (having lower scores than the mean) and those that fall to the right (having higher scores than the mean) are equal in number for a given number of units (or distance) from the midpoint in either direction. Also, in the normal distribution the number of cases having scores very different from the mean, median, or mode score drops off rather sharply. Most respondents are not strikingly different from the average respondent

ILLUSTRATION 13.3 Selected frequency distributions, high school survey, Middletown III Project

VARIABLE	NUMBER OF CASES (N)	PERCENT
<i>Age</i>		
11	1	.1
13	1	.1
14	177	10.9
15	472	28.6
16	458	27.8
17	378	22.9
18	145	8.8
19	17	1.0
Total	1649	100.2
<i>School Class</i>		
Freshman	498	29.9
Sophomore	463	27.8
Junior	406	24.4
Senior	296	17.8
Total	1663	99.9
<i>Schoolwork Time</i>		
None at all	127	7.7
Less than one-half hour	397	24.0
About one hour	519	31.4
Between one and two hours	444	26.8
More than two hours	167	10.1
Total	1654	100.0
<i>School Track</i>		
General	364	22.5
Business	292	18.0
College prep	416	25.7
Vocational	66	4.1
College prep & another	17	1.0
Business and vocational	2	.1
Other combinations	11	.7
Don't know	453	27.9
Total	1621	100.0
<i>Grades</i>		
A's or mostly A's	279	16.9
A's and B's	371	22.5
Mostly B's	240	14.6
B's and C's	403	24.4
Mostly C's	172	10.4
C's and D's	147	8.9
Mostly D's or below	37	2.2
Total	1649	99.9

ILLUSTRATION 13.4 Standard distribution compared with distributions for student age, work time, and grades.



in whatever characteristic is normally distributed. The greater the number of units of difference from that mean score, the less likely it is that any respondents will exhibit the deviant score, although the number deviating in the positive (more than the average or mean) direction is the same as the number deviating in the negative (less than the average or mean) direction if a distribution is "normal."

Now consider the frequency distribution for age, student work time, and grades charted in Illustration 13.4. Age has a distribution that resembles the standard normal distribution fairly closely, although it is skewed to the lower ages; that is, although there are nine age categories represented in the chart—11 through 19—if we ignore the statistically insignificant (one student for age 11 and 13, and none for age 12) category of students under age 14, more students are clustered at the early than at the later ages. Thus, students aged 14 or 15 account for 40 percent of the students, compared with only 10 percent at the two categories at the other end of the scale, ages 18 and 19. This kind of clustering toward one end of a range is sometimes referred to as a **skewed distribution**. The age distribution for Middletown high school students is skewed to the left, or more heavily represented in the lower than the upper age categories.

A more normal distribution fits student reports of time spent doing

homework each night. The item read, "How much time do you spend on schoolwork outside of class each day?" As can be seen in Illustration 13.3, the precoded answers ranged from "None at all" to "More than two hours." Illustration 13.4 shows how the students distributed themselves almost normally around the modal response of one hour (reported by 31 percent of the students), with approximately the same percentage in both the next higher and next lower categories (27 percent and 24 percent, respectively). In fact, one way to judge whether a distribution approximates the standard normal distribution is to see whether the mean and median are approximately the same as the mode. For the variable "Schoolwork Time," if the five categories are scored from 1 ("None at all") to 5 ("More than two hours"), the mode is 3.0, the mean 3.1, and the median 3.1. Student grades demonstrate a **bimodal distribution** with larger numbers of students reporting they receive mostly A's and B's or mostly B's and C's.

Levels of Measurement in Descriptive Analysis

The alert reader will have noticed that we stipulated "if the five categories are scored . . ." rather than saying something like "when the hours of study reported by the students were averaged," because in fact we did not find out how many hours the students studied, but only which of the five categories of study time best described their own situation. The categories themselves are of unequal size (or duration): the first, "None at all," represents zero minutes of study time; the second, "Less than one-half hour," represents between 1 and 29 minutes of study time; the third, "About one hour," may include students who studied 31 minutes (and thus did not fit the second category), up to students who studied 65 or 70 minutes. Presumably students who usually studied as much as 15 minutes more than one hour would have selected "Between one and two hours" rather than "About one hour" as the category that fit them. The final category, "More than two hours," is open-ended, perhaps ranging from slightly over two hours to as much as six hours of study per evening.

Thus, the amount of time represented in each category ranges from as little as one minute in category 2 to as much as six hours in category 5. While we do not know precisely where the respondents fall within the category limits, we do know that students in category 2 study more than those in category 1, those in category 3 more than those in category 2, and so on. In other words, we have here in the "Schoolwork Time" variable an ordering or ranking of responses, from "none," to "a little" to "a lot" of time studying. Because the intervals of time represented in the categories are of unequal duration, this set of categories cannot be considered an **interval scale**. By definition, an interval scale represents units of the same length or size.

What we have in this "Schoolwork Time" variable is known as **ordinal**

measurement, in which the researcher knows that a given category is greater or less than an adjoining category but the exact amount of difference is not known. Because the size of the intervals represented by each category is known to vary, it is not appropriate to add or subtract them, or to perform any of the common arithmetic computations with ordinal numbers. Thus, our reporting of a mean and a median for the "Schoolwork Time" distribution was technically incorrect, having meaning only if the categories were assumed to represent equal interval measurement.

The variable "Age" is easier to deal with because age is usually measured on an interval scale. Thus, a person 16 years old is said to be twice as old as an 8-year-old and half as old as a 32-year-old. Each year-interval is assumed to have equivalent length, and so arithmetic calculations such as computing means and medians are appropriate.

Let us illustrate another level of measurement by referring to the third panel of Illustration 13.3, "School Track." In coding of this variable, numbers were assigned to each track or combination of tracks, with 1 for "General," 2 for "Business," 3 for "College prep," and so on. However, these numbers do not refer either to a set interval or to any other kind of order. They are simply convenient ways of naming categories to make them readily usable in computer analysis. It makes no sense to perform arithmetic computations with them, for "General" has no numerical relationship to "Business," although the latter was punched 2 and the former 1. This assignment of identification numbers is known as **nominal measurement**.

Now examine the fourth panel of Illustration 13.3. Here again are precoded categories, this time for students' grades. Because the categories are defined only roughly, in terms of "mostly" or clusterings of adjoining grade levels in ambiguous combinations (for example, a student with 4 A's and 2 B's might have chosen the second category, "A's and B's," and so might a student with 2 A's and 4 B's), the exact size of each interval in this scale of grades is not known. However, the order is clear: "A's or mostly A's" ranks higher than "A's and B's" and the latter ranks higher than "Mostly B's," and so on. Thus, the "Grades" variable is another example of ordinal measurement. We have charted the distribution of student responses on this variable to illustrate another kind of distribution, namely the *bimodal* distribution. We have said that it is not legitimate to compute means or medians for a variable measured at the ordinal level; however, a mode, which simply refers to the most common response, is appropriate at every level of measurement, even the nominal. Note in the illustration that the distribution of grades does not reflect either the standard normal curve or the gradually descending linear curve represented by the "School Class" variable. Instead, there are two peaks or modes; one for the category "A's and B's" and another for the category "B's and C's." These two categories account for 47 percent of all respondents, leaving the remaining 53 percent to be distributed among the other five possibilities

in the seven-point scale. Note also in Illustration 13.4 that the distribution of grades is skewed toward the left, in this case toward high grades, such that 54 percent of the respondents are included in the first three categories (students who said they receive "Mostly B's" or above) compared with only 21 percent in the lower three categories (students who answered "Mostly C's" or below).

One more point about the interval level of measurement needs to be made. Some analysts distinguish between ordinary *interval scales*, whose only distinguishing characteristic is that each unit of measurement is the same size as every other unit of measurement, and **ratio scales**, or interval scales that include an absolute zero. Let us illustrate with reference to measuring temperature. Thermometers are calibrated with interval scales; each degree of Fahrenheit or Centigrade is the same size as every other degree, and so the difference between 35 degrees and 30 degrees is the same as that between 55 and 50 degrees. However, thermometers do not have an absolute zero point at which there is no temperature at all. Scientists have established the theoretical possibility of a point "Absolute zero," at which there is no temperature (at which those particles whose motion constitutes heat are at rest), and the hypothetical zero point is -273 degrees centigrade. Because in the case of temperature the absolute zero point is hypothetical, temperature is technically an interval, not a ratio scale.

On the other hand, the variable "Age" may be conceived as a ratio scale, with the moment of birth being defined as an age of zero. Similarly, if the "Schoolwork Time" variable discussed above had been measured in interval units of minutes or hours rather than in the five precoded ordinal categories, the variable would have reflected measurement at the ratio level, with students who spent no time at all studying at home at the absolute zero point.

Frequency Distributions and Comparisons

Looking at the frequency distributions of individual variables is essential even when the main objective of the project is testing relationships between two or more variables. One reason that such inspection is necessary is that some of the statistical tests used to measure the **statistical significance** of relationships between variables (i.e., whether an apparent relationship is "real" or probably due to chance) make assumptions about the way scores on a variable are distributed. Whether the distribution of scores on a given variable approximates the standard normal distribution may determine whether the use of certain statistical tests is legitimate.

A second reason for studying individual frequency distributions is that one frequently has to combine many categories into a few. For example, although one may have coded identification numbers for each reported

occupation, the research problem may be whether the working class differs from the business class. It is necessary to inspect individual occupations in order to make decisions about how to combine categories or recode the data to fit the working-class/business-class distinction. Also, some categories may have so few respondents that it makes sense to combine them with adjoining categories to produce enough cases to make it meaningful to cross-tabulate that variable by some other variable.

Sometimes examining the frequency distribution for a single variable is itself sufficient to satisfy research objectives. Knowing the **range** and distribution of family income in a city may be sufficient documentation of the size of the community's poverty class for administration purposes such as revenue-sharing or welfare program planning. A distribution of educational attainment among adults may identify the number and percentage of a neighborhood's adults who are potential clients for a high school degree equivalency program.

However, usually analysis of frequency distributions for single variables involves making comparisons between the obtained distribution and other distributions for the same variable. Such comparison is essential because without it, there is no way to ascribe meaning to the distribution revealed. Generally the descriptive analysis by single variables involves comparing the way a characteristic is distributed in the study population with the way it was distributed in the past or the way it is distributed in another population either presently or in the past. Only by comparison to such real or ideal distributions can one define what constitutes a "high" or a "low" score. To arbitrarily say that scores at the upper end of a distribution are "high" and those at the lower end are "low" has little meaning unless there has been such a comparison. For example, one might examine recent crime statistics and conclude that the frequency of crime is very high. However, trend comparisons of data from two points in time might reveal that crime rates have been going down. Similarly, comparisons with data from other societies might show that the rates are relatively low. The point is that the distribution of a characteristic (variable) is not meaningful in itself but takes on meaning only by comparison with another distribution. If there is no such reference point, people can erroneously draw any conclusion they want from a single frequency distribution or percentage based on such a distribution.

Being able to present a descriptive statement numerically may suggest that one "understands" a phenomenon, when in fact the numbers by themselves are uninterpretable. For example, to say that 26 percent of the Middletown high school students in 1977 were enrolled in the college preparatory track may be interpreted as a tragic situation, a favorable situation, or a condition that merits no action whatever. Conceivably, 26 percent of students in every public high school in the nation might be enrolled in a college preparatory track. Only when that figure is compared

with an ideal (e.g., the school board has set as policy that *half* of all students should be enrolled in college prep track) or a literal benchmark (e.g., in 1947 only 10 percent of Middletown students were in the college track) does the descriptive statistic take on meaning.

Thus, the key to effective descriptive analysis is comparison, and the choice of comparison groups or other ideal standards for comparison may determine the utility of the descriptive analysis. Our own study of Middletown high school students involved much comparative analysis over time, generally comparing the distribution of students along a single variable in 1924–25 with the distribution along that same variable in 1976–78. Let us illustrate by referring to some items on religious belief that Robert and Helen Lynd (1929) included in their survey of Middletown high school students in 1924.

After recomputing Lynd’s published percentages to exclude “uncertain,” which was a possible response in their survey but not in ours, and grouping the four precoded responses in our survey (strongly agree, agree, disagree, and strongly disagree) into two, to match their format, frequency distributions for each item amounted to two numbers, those who agreed with the item, and those who disagreed. In our presenting the data, only one of these possibilities had to be reported, since the combination of “agrees” and “disagrees” always added up to 100 percent. Illustration 13.5 presents the percentage answering “True” or agreeing with each of four statements about religion in 1924 and 1977.

Each percentage in Illustration 13.5 is one of two categories of a frequency distribution for an item, the other category being the “false” or “disagree” responses not shown. It is possible to describe the percentage

ILLUSTRATION 13.5 Percentage of Middletown High School students agreeing with various statements about religion, 1924 and 1977.

ITEM	PERCENT ANSWERING “TRUE” OR AGREEING	
	1924	1977
Christianity is the one true religion and all peoples should be converted to it.	94% (521)	38% (886)
Jesus Christ was different from every other man who ever lived in being entirely perfect.	83 (526)	68 (564)
The purpose of religion is to prepare people for the hereafter.	60 (493)	53 (581)
It is wrong to go to the movies on Sunday.	33 (556)	6 (594)

*Figures in parentheses are percentage bases (N’s).
Source: Middletown III high school survey. For further detail on the methods of both surveys and other topics of comparison, see Caplow and Bahr (1979).

distributions for either year. Thus, it is accurate to say that almost everyone in 1924 believed that Christianity was the one true religion, and that one-third of the students believed that it was wrong to go to the movies on Sunday. However, when the 1924 distributions are used to "anchor" the 1977 distributions, a higher order of interpretation is possible. Then it is possible to say that 38 percent agreement to the "Christianity is the one true religion" statement represents a fairly low rate of agreement. With the 1924 benchmark it is apparent that the 1977 students are much more tolerant of other religions than were their 1924 counterparts.

Observe also that for two of the items—"Jesus Christ was different . . ." and "The purpose of religion . . ."—15 percentage points or less separate the distributions of the 1924 and 1977 students. In other words, there has been some decline, but not much. Again, the 1924 benchmark figures allow the analyst to draw some conclusions that would be impossible with only a single set of figures.

Our use of the 1924 Middletown data as a comparison point to interpret the 1977 figures does, in fact, introduce a second variable into the frequency distributions: time, or history. What Illustration 13.5 does, in effect, is to cross-tabulate a set of four single variable distributions from 1924 with the same four from 1977. Indeed, any use of a benchmark distribution, ideal or actual, whether representing a previous condition of the same population or a contemporary distribution from a different population, technically moves us to the analytical strategy of cross-tabulation.

If we are forced to use only a single distribution, then descriptive analysis does us little good other than to aid in collapsing categories or deciding whether a particular distribution is close enough to a normal distribution to justify the use of certain statistical techniques in subsequent analysis. Looking at the univariate distributions is usually only a preliminary step. Most social analysis involves cross-tabulation of some sort, and we will now turn to that topic.

V. CROSS-TABULATION AND CONTINGENCY TABLES

The term **cross-tabulation** refers to the combining of the frequency distributions (generally collapsed into a manageable number of categories) on two or more variables. In this phase of analysis, the researcher is asking questions like these: How is the distribution of responses on my dependent variable (for example, attitudes about religion) affected by the distribution of responses on an independent variable (say, gender, or years of schooling)? Or the analyst may be concerned with the way the distribution of responses on a dependent variable is influenced by various combinations of independent variables.

In the simplest kind of cross-tabulation, a *two-by-two table*, the rela-

tionship between two variables, each measured in only two categories, is shown. Illustration 13.6 portrays a two-by-two table in which responses from the 1977 Middletown High School survey (freshmen and blacks excluded, to conform to the Lynds' sample) are divided into "Agree" and "Disagree" for the item "It is entirely the fault of a man himself if he does not succeed" and are cross-tabulated by gender.

Note that the figures in the illustration are absolute numbers of cases, not percentages. Analyzing the relationships in such a table may involve computation of percentages based on the *row totals* (that is, what percentage of students who agreed were males), *column totals* (what percentage of males agreed with the statement), or the *grand total* of all students represented in the table (what percentage of all students were males who agreed, what percentage were females who agreed, and so on).

The distribution of students in Illustration 13.6 does in fact reveal a fairly strong relationship between gender and support for the statement that a man is responsible for his own success. Males were more likely to agree than females (53 percent versus 41 percent, respectively). That 12-percentage-point difference, for a sample as large as this one, suggests that gender really does have something to do with agreement on the individualistic ethic reflected in the item.

While the simplest form of cross-tabulation occurs in the two-by-two table, more complex tables can also be created. For example, if the researcher is interested in looking at relationships among three variables, two tables can be created, each showing a relationship between two variables while controlling for the third.

Creating cross-tabulations or **contingency tables** such as that illustrated is typically the first step in assessing the type of relationships that exists between variables. The researcher can examine the visual presen-

ILLUSTRATION 13.6 Cross tabulation of attitudes about success by gender, Middletown High School students, * 1977.

ITEM		GENDER		
		Males	Females	Total
It is entirely the fault of a man himself if he does not succeed.	Agree	227	196	423
	Disagree	200	286	486
	Total	427	482	909

* Freshmen and black students were omitted from these tabulations to make the sample conform to that of the Lynds.

tation of the data, percentages can be calculated and compared for the columns and rows in the table, and so on.

VI. TESTING RELATIONSHIPS BETWEEN VARIABLES

We have discussed ways of presenting data when two or more attributes or characteristics (often referred to as independent and dependent variables) are being described. However, in addition to showing the number of respondents having specific characteristics, the researcher will frequently want to report additional information. For example, is the relationship that is observed between variables a significant one? How much of the variation that is observed in the dependent variable can be explained by the independent variable or some combination of independent variables? To answer questions such as these, the researcher will typically employ a variety of statistical techniques.

Means, modes, medians, and percentages, which we have discussed, are generally referred to as **descriptive statistics**: they help the researcher describe his or her findings. Statistics that allow the researcher to determine whether or not a relationship is statistically significant (basically, an assessment of whether it is "real" or could have occurred by chance) are usually referred to as **inferential statistics** or **tests of significance**. Statistics that assess the strength of the relationship between variables or the amount of variation in the dependent variable that can be explained by the independent variable are referred to as **measures of association**.

In the analysis of data, the researcher will typically use both tests of significance and measures of association. **Statistical significance** should not be confused with **substantive significance**, that is, with what is important from a practical or a theoretical point of view. As we will note, the statistical significance of an observed relationship between variables is affected both by the strength of the relationship and by other factors such as sample size. Therefore, one might obtain a statistically significant relationship that is so weak as to be trivial. As Blalock (1960: 126) has noted: "Statistical significance can tell us only that certain sample differences would not occur very frequently by chance if there were no differences whatsoever in the population. It tells us nothing directly about the magnitude or importance of these differences." The strength or importance is assessed using measures of association which will be discussed later.

The particular test of significance or measure of association one chooses will be determined in part by whether the level of measurement one obtains with the data is nominal, ordinal, or interval. We will discuss both tests of significance and measures of association that can be used with each level of data.

Tests of Significance

The purpose of using tests of significance is to be able to say something about the characteristics (**parameters**) of a population from observations (statistics) made on a sample of respondents. In discussing significance tests, we establish a direct link with Chapter 3, on sampling. When we collect data from a sample, we need to ask whether the findings can be generalized to the larger population or are simply a function of the particular sample studied.

Thus the **probability** that what was observed occurred by chance should be determined. It is assumed that random sampling procedures have been properly applied. A significance test will produce a number (called the **significance level**) that tells us how confident we can be that the findings in the sample are applicable to the population. Typically, social scientists use the .05 or .01 level of significance, which means that if 100 samples were selected, in 95 or 99 of them the difference observed would appear.

Statistical tests of significance are a critical aspect of hypothesis testing. Frequently the researcher will hypothesize a relationship between two or more variables (for example, between educational level and religious behavior) and then will use tests of significance to determine whether or not such a relationship actually exists. Hypotheses are usually stated in the *null* form. That is, our hypothesis may be that there is no relationship between educational level and religious behavior. We apply our statistical test to see if we can reject the **null hypothesis**, to demonstrate that there is indeed a relationship between variables.

Statistical tests of significance are based on a number of rather stringent assumptions. The reader should consult a beginning statistics text for more detailed discussions of these. For illustrative purposes, we will briefly identify tests that can be used for different levels of data.

Nominal and ordinal data The test of significance that is most commonly used for both nominal and ordinal levels data is the **Chi-square** (χ^2). In calculating χ^2 the researcher makes a determination of the difference between the *observed* frequencies that occur in a cross-tabulation and the frequencies that one would *expect* assuming no relationship between the variables. We refer to the latter as *expected frequencies* since they are what would occur if the two variables were indeed independent of each other. If the observed frequencies are sufficiently different from the expected frequencies, then we conclude that the differences would not have occurred by chance and that there is a real relationship between the variables.

The formula for calculating Chi-square is as follows:

$$\chi^2 = \frac{\sum(O - E)^2}{E}$$

O refers to the observed value and E the expected value. Expected frequencies are calculated by multiplying the row total by the column total for any given cell and then dividing it by the total sample N. This can be shown by Illustration 13.7.

For this case, the expected frequencies for the four cells in the table would be calculated as follows:

$$E_{(a)} = \frac{(30 \times 35)}{60} = 17.5$$
$$E_{(c)} = \frac{(30 \times 25)}{60} = 12.5$$

$$E_{(b)} = \frac{(30 \times 35)}{60} = 17.5$$
$$E_{(d)} = \frac{(30 \times 25)}{60} = 12.5$$

x² would then be computed as follows:

$$x^2 = \frac{(7.5)^2}{17.5} + \frac{(7.5)^2}{17.5} + \frac{(7.5)^2}{12.5} + \frac{(7.5)^2}{12.5} = 3.21 + 3.21 + 4.50 + 4.50 = 15.42$$

To determine whether or not Chi-square is significant, we must calculate the degrees of freedom. This is determined for contingency tables as follows: Degrees of Freedom = (R - 1) × (C - 1) where R = number of rows in the table and C = number of columns. For the above case, the number of degrees of freedom is (2 - 1) × (2 - 1) = 1. A Chi-square table listed in most statistics texts would tell us that a Chi-square of 15.42 with one degree of freedom is significant at the .001 level. That is, we can be confident that the relationship we have observed is real and would occur by chance less than one time in a thousand.

Numerous other tests are available for ordinal level data; the researcher should use the one that best fits the needs of the research problem.

ILLUSTRATION 13.7 Relationship between political party and income level.

POLITICAL PARTY	INCOME LEVEL		
	High	Low	Total
Republican	25 (a)	10 (b)	35
Democrat	5 (c)	20 (d)	25
Total	30	30	60

Interval data or data with different levels of measurement If the researcher has a data set in which one of the variables is nominal and the other is interval, a variety of difference in means tests can be used to determine significance of relationships. For example, if the researcher has two independent random samples, a **difference of means** or **difference of proportions** tests can be calculated. In a difference of means test, the researcher simply makes a comparison between the means for two independently selected samples in order to determine whether the differences in the two samples are sufficiently large that they would not have occurred by chance. One works directly with the two sample means to compute a coefficient that assesses the likelihood that an observed difference between the two samples is real. Difference of proportions tests are simply a special case of testing the difference between means. In this instance, the values are translated into proportions and one assesses the magnitude of the difference that occurs between the proportions.

To illustrate the former, suppose the researcher has drawn a sample of respondents from counties that are predominantly urban and another sample from counties that are predominantly rural and is interested in determining whether the level of educational attainment is different for the two samples. A difference of means test could be calculated to determine whether any difference observed is statistically significant.

If multiple nominal categories are present and the dependent variable is measured intervally, **analysis of variance** can be used. Basically, analysis of variance represents an extension of the difference of means test by allowing the researcher to assess the differences among the means of more than two samples simultaneously. In computing analysis of variance tests, the researcher works directly with variances (deviations from the means) rather than with means. Analysis of variance procedures identify the amount of total variance that is attributable to each of the separate samples and to their various combinations. This latter characteristic is referred to as testing for *interaction* effects; that is, different levels and combinations of the independent variables can be examined to determine their effect on the dependent variable.

If **correlation analysis** is used, the most direct procedure is simply to test whether or not the simple r^2 or the multiple R^2 is statistically significant. Here the researcher is simply asking whether a correlation of the size obtained would have occurred if there was not a relationship between the variables in the population being studied. Detailed discussions of all of these techniques are available in most statistics texts.

Measures of Association

The analytical statistics used to determine the strength of a relationship between variables are generally labeled *measures of association*. These allow the researcher to determine the amount of change in one variable

that is a function of change in another variable or set of variables. For example, we might hypothesize that increased levels of educational attainment have a negative effect on one's religious beliefs and behavior. Several studies (see Caplovitz and Sherrow, 1977, for example) have shown that there is an important strain between commitment to intellectual pursuits and commitment to religion and that the former tends to undermine the latter. On the basis of such previous work, we would predict a negative association between religion and education, that is, the higher one's educational attainment, the lower one's religious observance. Illustration 13.8 presents data collected by the authors that examine this relationship for a large sample of members of the Mormon Church.

If we apply measures of association to the data presented in the illustration, we are basically asking how much variation in religious observance can be explained by variation in educational attainment. Education would be considered the independent variable and religious observance the dependent variable. If religious observance can be predicted perfectly from educational attainment, then we have a very strong relationship—our measure of association would take on the value of 1.0, indicating that a change in educational attainment automatically produces a change in religiosity. If, on the other hand, religious observance is unaffected by educational level, then our measure of association would approach the value of 0.0. In most social science problems the relationships observed between independent and dependent variables are neither perfectly associated or perfectly independent. In other words, knowing the score a respondent receives on an independent variable may improve prediction of scores on the dependent variable over what could be done without such information, but the improvement in prediction may not amount to much.

As was noted, the measure of association that one selects for testing

ILLUSTRATION 13.8 Relationship between educational level and attendance at religious services, Mormon sample.

EDUCATIONAL LEVEL	ATTENDANCE AT RELIGIOUS SERVICES	
	Less than Weekly	Weekly
Grade school	62	32
Some high school	94	87
High school graduate	353	267
Some college	217	403
College graduate	70	171
Graduate school	75	302
N	871	1,262

the strength of a relationship is determined by the level of measurement that one has attained—nominal, ordinal, interval, or ratio. Measures of association that are applicable when one's data are either nominal or ordinal are usually referred to as **coefficients of association**. When interval or ratio level data are available, the measures are typically called **correlation coefficients**.

The most frequently used measures of association exhibit an important property that allows them to be interpreted in terms of the *reduction of error in prediction* that is achieved by knowledge of the independent variable. Costner (1965) refers to this property as **proportional reduction in error (PRE)**. To return to our illustration on religious behavior and educational achievement, a PRE measure would allow us to indicate the amount of reduction in error that could be achieved in predicting one's religious observance given knowledge of educational level over what could be achieved without that knowledge. If there is a strong relationship between the variables, knowing a respondent's education greatly reduces the researcher's error in predicting the respondent's religious behavior. If, on the other hand, the variables are virtually unrelated, then knowledge of educational attainment does not produce any substantial reduction in error in the prediction of religiosity.

All measures of association having proportional reduction in error properties share a common logic (Mueller, Schuessler, and Costner, 1970) which focuses on the amount of reduction in error in predicting the dependent variable that can be achieved by knowledge of the independent variable. The prediction rules for each measure are different, but all contain the following elements (Mueller, Schuessler, and Costner, 1970: 247–248): (1) a rule for predicting some characteristic of a dependent variable from knowledge of its own distribution; (2) a rule for predicting the same characteristic of the dependent variable from some characteristic of the independent variable; (3) a definition of what constitutes error and how it shall be measured; and (4) a definition of the measure of association that takes the form:

$$\text{Proportional Reduction in Error} = \frac{\text{Error by Rule (a)} - \text{Error by Rule (b)}}{\text{Error by Rule (a)}}$$

Nominal data Although several measures of association are available for each level of measurement, we will select just one for each level for illustrative purposes. If one's data have only nominal properties, one can test the strength of the relationship by using **Lambda**. Lambda is based on the assumption that if one has nominal data, the only measure of central tendency that makes sense is the mode. If the mode is used, we can make predictions about variation in our dependent variable without additional knowledge of our independent variable by predicting all cases will fall at the mode. An error will occur in prediction when any case does not fall at the mode. This can be demonstrated by Illustration 13.9. In this example

ILLUSTRATION 13.9 Relationship between parent and child political party identification.

PARENTS' POLITICAL PARTY	CHILD'S POLITICAL PARTY	
	Democrat	Republican
Both Democrats	41	12
One Democrat and one Republican	16	14
Both Republicans	14	21

we are interested in the strength of the relationship between parental political party identification (the independent variable) and the party identification of their children (the dependent variable). Remember that our rule for estimation is the mode, and our definition of "error" is anything that does not fall at the mode. In the example, if we did not know the party identification of parents and we used the mode as our prediction rule, we would predict that all children would be Democrats. Our prediction error would be 47/118 or .398. That is, if we predicted that all children were Democrats, we would be wrong almost 40 percent of the time.

Now suppose we add the information on parental political party. Here we would use modal categories for each *row* in the table; in other words, if we know that both parents are Democrats, we would predict that the child is a Democrat; if we know that one parent is a Republican and one is a Democrat, we would predict that the child is a Democrat; and if we know that both parents are Republicans, we would predict that the child is a Republican. The predictions in the first two cases are the same as those we would make using the mode rule. Therefore, any reduction in error in prediction occurs in the third case: we predict that the children of Republican parents will be Republicans. The formula for calculating Lambda is as follows:

$$\lambda_b = \frac{\left(1 - \frac{f \cdot m}{N}\right) - \left(1 - \frac{fam}{N}\right)}{1 - \frac{f \cdot m}{N}} = \frac{\sum \frac{fam}{N} - \frac{f \cdot m}{N}}{1 - \frac{f \cdot m}{N}}$$

$\left(1 - \frac{f \cdot m}{N}\right) =$ error in prediction without considering (a)

$\left(1 - \frac{fam}{N}\right) =$ error in prediction considering (a)

Here we assume that (b) represents the dependent variable (child's party identification) and (a) represents the independent variable (political party identification of the parents). We are concerned with the reduction in prediction error for (b) that comes from a knowledge of (a). The Lambda for the table is calculated as follows:

$$\begin{aligned}\Sigma \text{ fam} &= 41 + 16 + 21 = 78 \\ f \cdot m &= 71 \\ \lambda_b &= \frac{78 - 71}{118 - 71} = .149\end{aligned}$$

This tells us that error in predicting the child's political party can be reduced by about 15 percent given the knowledge of parents' political party over what could be achieved without such information about the parent.

Several problems limit the usefulness of Lambda. For example, if all of the modal responses fall in the same column, the value of Lambda will be zero. Therefore, the researcher must consider the properties of alternative measures and adopt those most applicable to the data. Further detail on the comparative advantages and disadvantages of alternative measures is available in most introductory statistics texts.

Ordinal data Probably the most widely used ordinal measure of association is **Gamma**. We can illustrate the use of Gamma by referring to the data presented earlier on the relationship between education and religious behavior. The variables both have ordinal properties in that respondents can be ranked in terms of the number of years of school completed and the frequency of attendance at religious services.

Because Gamma considers the ordinal properties of variables, its basic prediction rule is based on the relationship between pairs of observations. Basically, the researcher is interested in whether randomly drawn pairs from a sample exhibit consistent (concordant) or inconsistent (discordant) relationships with each other. Gamma is calculated by determining the number of concordant pairs in a table relative to the number of discordant pairs. The prediction rule is as follows:

$$\text{Gamma } (\gamma) = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{(\text{number of concordant pairs}) + (\text{number of discordant pairs})}$$

To determine the number of concordant pairs, each cell is multiplied by the sum of those cells to right and below. The number of discordant pairs is determined by multiplying each cell by the sum of those to the left and below. Using this procedure, one would calculate the relationship between educational level and religious observance (Illustration 13.8) as follows:

number of

$$\text{concordant pairs} = 62(87 + 267 + 403 + 171 + 302) + 94(267 + 403 + 171 + 302) + 353(403 + 171 + 302) + 217(171 + 302) + 70(302) = 616,711$$

number of

$$\text{discordant pairs} = 32(94 + 353 + 217 + 70 + 75) + 87(353 + 217 + 70 + 75) + 267(217 + 70 + 75) + 403(70 + 75) + 171(75) = 256,007$$

$$\text{Gamma} = \frac{616,711 - 256,007}{616,711 + 256,007} = \frac{360,704}{872,718} = .41$$

Our Gamma of .41 indicates that error in prediction of religious attendance can be decreased by just over 40 percent if we know a person's educational attainment. In other words, for the sample of Mormons whose educational attainment and church attendance are cross-tabulated in Illustration 13.8, there is a strong positive relation between education and attendance. This means that better educated individuals tend to be the most frequent church attenders, and the less educated attend more infrequently. The size of the Gamma indicated a fairly substantial improvement in the prediction of the dependent variable, church attendance, given the knowledge of rank on the independent variable, education.

There are many measures of association available for ordinal data, and again the reader is referred to an introductory statistics text for discussions of other measures and their advantages and disadvantages.

Interval data If we are able to obtain interval level measurement with both variables or with our dependent variable, we can use **correlation and regression analysis**. These techniques are perhaps the most widely used statistical procedures in the social sciences and can be used to test relationships between a dependent variable and a single independent variable or several independent variables. Correlation and regression analysis also allows the researcher to test for the impact of a single independent variable on the dependent variable while statistically controlling for the effects of other independent variables. Because of the importance of these procedures, we will briefly discuss simple, multiple, and partial correlation and regression techniques.

SIMPLE CORRELATION AND REGRESSION TECHNIQUES When we are interested in the relationship between only two variables or when we want to predict some value for one variable from knowledge of the value of one other, we speak of **simple linear regression**. Simple linear regression is used to determine the degree to which one variable changes with a given change in another variable. After we determine how much a dependent

variable, we can predict what the value of other observations of the dependent variable will be from simply knowing the value of the independent variable.

The simple linear regression equation $Y = a + bX$ tells us that every time X increases one unit, Y increases b times one unit. Once the values for the constants (a and b) in the equation are determined, the value of Y can be predicted for any given value of X .

The equation $Y = a + bX$ provides a straight line which best describes the relationship between X and Y . The constant (a) represents the point at which the line crosses the Y axis (i.e., where $X = 0$) and the constant (b) represents the slope of the line, that is, the magnitude of the change in Y for a given change in X . The steeper the slope, the larger the change in Y for each change in X . If the values of Y are plotted against X in a scatterplot, this regression equation will determine the line that will best describe the relationship. The greater the dependence between X and Y , the nearer their plotted values will come to falling directly on the regression line.

In order to be able mathematically to examine the predictability of Y from X , the regression line we need is that one that will best represent *all* of the observations that have been made. In other words, a compromise regression line must be determined which will come as near as possible to agreeing with all observations even though it may not exactly agree with any of them. The method of "least squares" has been worked out by mathematicians as the means of obtaining such a line. The formula given above ($Y = a + bX$) is this mathematically determined line. This is also referred to as the *line of best fit*, since for each value of X , values of Y come as near as possible to agreeing with all different Y values observed.

A simple extension of this allows prediction of the value of any Y from knowledge of X . However, we should stress that an estimating equation of this type can be used only within the range covered by the original observations. For example, if we were studying the relationship between income and attitudes toward social welfare and our observations covered a range of income from \$3,000 to \$10,000, we would be on very questionable ground in predicting attitudes from later observations of individuals whose range of income included figures less than \$3,000 or more than \$10,000.

Simple correlation is concerned with measuring the degree of relationship between two variables. The correlation coefficient (r) is a single number that reflects how well the linear or other equation obtained in regression analysis describes or explains the relationship between two variables. As Blalock notes, sociologists are often interested in discovering which of a series of variables are most closely related to a given dependent variable (Blalock, 1960: 285). If this is the case, prediction and regression analysis are of secondary importance.

Correlation is actually very closely related to regression. The coeffi-

cient (r) is simply a measure of the deviance of the observation from the line computed by the least squares method in regression analysis. Thus (r) is simply a ratio of the variance in Y accounted for by X , to the total variance in Y .

$$r = \frac{y'}{y}$$

In other words, when the values of Y are estimated from the values of X according to the previous straight-line equation, then the proportion of the variation in Y so accounted for is (r). The coefficient is simply a measure of how large the variation in the estimated values is in proportion to the variation in original values.

The major problem with (r) is that it is difficult to interpret. In reporting results, therefore, we usually report (r^2), which is referred to as the **coefficient of determination**. The coefficient of determination may be said to measure the percentage of variance in Y determined by X , or simply, "the **explained variance**."

When do we use correlation and when regression? The answer to this question is given by Blalock as follows:

When interest is focused primarily on the task of finding out which variables are related to a given variable, we are likely to be mainly interested in measures of degree or strength of relationship such as correlation coefficients. On the other hand, once we have found the significant variables we are more likely to turn our attention to regression analysis in which we attempt to predict the exact value of one variable from the other. (Blalock, 1960: 273)

The use of either procedure, however, generally assumes that both our dependent variable and independent variable(s) are measured on interval scales.

MULTIPLE CORRELATION AND REGRESSION ANALYSIS. Multiple correlation and regression analysis is nothing more than a straightforward extension of simple correlation and regression to include any number of interval scales, any one of which can be assumed to be the dependent variable and the remainder independent variables. Thus if our concern is with prediction, we talk about prediction of the dependent variable X_1 from independent variables $X_2, X_3, \dots X_n$. If our concern is with the degree of relationship, multiple correlation refers to the *total* variation in X_1 that can be explained by all of the independent variables acting together.

In many of the types of problems dealt with in social science research, variation in one variable may be expected to reflect change in a number of other variables, all acting at the same time. The physicist and the biologist can use laboratory methods to deal with problems of multiple causation.

Under laboratory conditions all of the variables except the one whose effect is being studied may be held constant, and the change that occurs in the dependent variable can then be attributed to that one remaining variable. However, for most of the problems that social scientists deal with, such controls are not possible. This is why multiple correlation and regression are such useful research tools. The effects of a complex series of factors can be considered simultaneously and their combined effects on the dependent variable determined.

Where we have a dependent variable influenced not by a single independent variable but by two or more independent variables, the relationship can be symbolically represented by the equation

$$X_1 = a + b_2X_2 + b_3X_3 + \dots b_nX_n + e$$

This equation is termed the **multiple regression** equation, meaning that X_1 is explained in terms of $X_2, X_3, \dots X_n$, respectively, excluding, or net of, the associated influences of other independent variables. The (e) term represents the effect of all other variables not included which actually are involved in any problem. The researcher selects what seem to be the most important independent variables (or the ones accessible to measurement) and assumes the influence of variables not included (e) to be relatively unimportant.

For the case of three variables, the regression equation is

$$X_1 = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3$$

As in simple regression, (a) and (b) are constants. The extension of the subscripts following the b's tell us that the first ($b_{12.3}$) is the regression coefficient of X_1 on X_2 keeping X_3 constant, and so on.

The computing formulas for the constants are as follows when only two independent variables are involved:

$$a_{1.23} = \bar{X}_1 - b_{12.3}\bar{X}_2 - b_{13.2}\bar{X}_3$$

$$b_{12.3} = \frac{b_{12} - (b_{13})(b_{32})}{1 - b_{23}b_{32}}$$

$$b_{13.2} = \frac{b_{13} - (b_{12})(b_{23})}{1 - b_{32}b_{23}}$$

The subscripts following each of the regression coefficients simply indicate the total regression of each of the variables against one other variable. These same formulas can, of course, be extended to include any number

of independent variables but the computations involved become quite complex if one tries to work with more than three or four variables.

Multiple correlation is concerned with measuring the total variation in a dependent variable that can be explained by a series of independent variables acting together. Correlation analysis is often used in the early stages of an investigation when a large number of variables are under scrutiny. As evidence accumulates, correlation techniques allow the researcher to discard some possible independent variables because others are found to be adequate to account for the observed variation in the dependent variable. In other words, variables that do not add to the amount of variance explained by an existing set of variables can be set aside.

Conceptually we can state that in every scientific problem there is 100% variation in the dependent variable that requires explanation. The multiple correlation coefficient (R) is a ratio of the variation that we have been able to account for with our independent variables to the total variation that exists. Symbolically, if we let Y equal the total variation and Y' equal the explained variation, then Y'/Y equals R^2 . To put it in different words, R^2 can be defined as equal to the sum of squares because of regression divided by the total sum of squares. R^2 , then, measures the proportion of total variation about the mean (X_i) explained by the regression. The difference between the two sums of squares is, of course, the error or unexplained variance.

These three sums are needed for testing the significance of the relationship through the use of analysis of variance.

Blalock (1960) diagrams the relationship between the multiple correlation coefficient and the simple or total coefficients:

$R^2_{1.23} =$	r^2_{12}	+	$r^2_{13.2}$	$(1 - r^2_{12})$
Proportion	Proportion		Additional	Proportion
explained by 2 and 3	explained by 2		proportion	unexplained by 2
			explained by 3	

When this is expanded just one term with the introduction of one more independent variable, the complexity increases greatly:

$$R^2_{1.234} = r^2_{12} + r^2_{13.2} (1 - r^2_{12}) + r^2_{14.23} [1 - r^2_{12} - r^2_{13.2} (1 - r^2_{12})]$$

It should be noted again that R^2 is used in reporting results since it is more easily interpreted. R^2 is a measure of explained variance or a percent of the total variation that is accounted for by the independent variables acting together. Thus, if R equals .50, we have been able to account for 25 percent of the variation.

Certain assumptions must be met if we desire to generalize from

sample correlation and regression coefficients to the larger universe from which the sample was drawn. The most important of these are as follows:

1. Random sampling must be assumed.
2. Each variable must be normally distributed in the population.
3. Variation in the dependent variable is the same regardless of the level of the independent variable(s). In other words, we must assume normality for the dependent variable for each fixed level of the independent variable. This is assured if each of the variables are normally distributed in the population and the sample is selected through random techniques.
4. Interval level measurement is required.

If these assumptions are met, inferences can be made from the results. Also sample correlation and regression coefficients become unbiased estimates of the population parameters and confidence intervals and probability statements can be made. If, on the other hand, we are not interested in generalizing from our results but only in discovering how given variables are related, correlation and regression can be used without concern for such questions as sampling procedures used.

Finally, let us suggest that correlation and regression analysis is best suited for problems where (1) you have a single dependent and one or more independent variables, (2) numerical data can be obtained for each variable on a continuous value scale (interval measurement), (3) the relationship can be represented by a linear straight line or simple form of a curve, and (4) the effect of each of the dependent variables is additive, that is, there is low intercorrelation among the independent variables.

The major limitations of correlation and regression analysis are (1) it does not establish cause-effect relationships, (2) it is often very difficult in social science research to meet the assumptions of correlation models, and (3) it is often hard to determine the extent and effect of intercorrelations between independent variables especially if there are many of them. It is especially important to keep in mind that correlation does not mean causation. One should always be alert for "nonsense" or spurious correlations.

PARTIAL CORRELATION AND BETA COEFFICIENTS Multiple correlation and regression techniques provide us with a means of determining the combined effect of a series of independent variables on a single dependent variable. In addition to having measures of the combined effect, it is often desirable to have measures of the *relative* importance of each of the individual independent variables taken separately, while simultaneously allowing for the variation that is associated with the other independent variable (Ezekial, 1941: 213). There are two different types of measures that will give us this information: the coefficient of partial correlation and the beta coefficient.

Coefficients of **partial correlation** reflect the relationship between the

dependent variable and each of the several independent variables, while eliminating any tendency for the remaining independent variables to affect the relationship. Thus, the partial correlation coefficient $r_{12.34}$ shows the correlation between X_1 and X_2 while controlling for the effect of X_3 and X_4 on the relationship. It may be defined, therefore, as a measure of the extent to which that part of the variation in the dependent variable that was not explained by the other independent variables can be explained by the addition of a new factor.

When it comes to comparing the relative effect of each of the net regression coefficients on the relationship, however, a major problem emerges. It will seldom be the case that each of the independent variables being used is stated in the same units. Thus one variable may be stated in dollars of income, another in years of education, another in number of children, and so on. As Blalock states, "If the partial b 's are to be used to compare various independent variables as to their relative abilities to produce changes in the dependent variable, we must correct for the fact that there will undoubtedly be differences in scales involved" (Blalock, 1960: 344).

The problem is solved by simply stating each of the net regression coefficients in standard units so they can be compared. The standardization is achieved by dividing the net coefficients by their **standard deviations**. These adjusted coefficients are termed *beta coefficients* or *beta weights*. The formulas for attaining the weights or the relative importance of each of the independent variables in the above regression equation are as follows:

$$B_{12.34} = b_{12.34} \frac{Sx_2}{Sx_1} \qquad B_{14.23} = b_{14.23} \frac{Sx_4}{Sx_1}$$

when

$$B_{13.24} = b_{13.24} \frac{Sx_3}{Sx_1} \qquad Sx_1 = \sqrt{\frac{\sum x^2_1}{n}} \quad \text{and so on.}$$

Both partial correlation coefficients and beta coefficients provide a measure of the relative importance of each of the independent variables on the relationship. They will not generally both be used with the same problem for they tell us somewhat different things. As Blalock notes, "The partial correlation is a measure of the amount of variation explained by one independent variable after the others have explained all they could. The beta weights, on the other hand, indicate how much change in the dependent variables is produced by a standardized change in one of the independent variables when the others are controlled" (Blalock, 1960: 345). However, it is almost always the case that the two coefficients end up ranking variables in the same order of importance.

VII. INDEXING AND SCALING

Up to this point in our discussion of analysis procedures we have treated our variables as if they were based on a single item or indicator. For example, educational level is measured by a single question that determines the number of years of school completed. Many if not most of the variables that are used by social scientists use multiple indicators; that is, several items are combined to represent a single variable. One of the primary reasons for doing this is to increase the reliability of one's measurement. If several items are used to measure a single variable it is assumed that the product will be more reliable than would any of the single measures used alone. For example, student exam scores are based on responses to a number of questions; it is assumed these are a better measure of ability and performance than response to a single question.

The concepts of **indexing** and **scaling** refer to the combination of several measures into a single, composite variable. This score is then treated as the variable of interest. For example, social scientists frequently talk about a community's "quality of life." Quality of life is a composite measure that summarizes a series of individual measures that might include such things as the community's crime rate, its infant mortality rate, its cost of living, the availability of shopping and public services, and other indicators of what goes on in a community that make it a more or less desirable place to live.

The most frequently used example of these procedures in the social sciences is found in the construction of attitude scales. Basically, attitude scales combine a series of items or questions into a single score that is then used as the measure of a respondent's attitude. This score is then submitted to statistical analysis using the procedures discussed above. In this section we will briefly review the major methods that are used for the construction of such attitude scales.

Thurstone Scales

Like single-item indicators, attitude scales have different measurement properties. One of the first attempts to develop attitude scales is found in the work of L. L. Thurstone (1928), who sought to develop scales that would have interval-level properties. Thurstone assumed that such scales could be analyzed by means of procedures that can be used with any other type of interval data. Whether or not Thurstone scales actually achieve interval measurement is still debated in the literature.

The most widely adopted method of attitude scaling developed by Thurstone is referred to as the **method of equal-appearing intervals**. Several steps are used in constructing such a scale. Typically, the researcher will assemble a large number of statements about the subject of interest.

For example, if a study is being done to measure attitudes toward abortion, the first step will involve developing a number of statements that deal with the topic of abortion. The statements are written to be brief, unambiguous, and directly related to the topic of interest. Second, the statements are presented to a number of individuals who are asked to evaluate each one in terms of its favorability or unfavorability toward the topic of interest. The judges are asked to sort the statements into eleven piles with one end representing those statements that are most favorable and the other those that are least favorable. Third, the researcher computes a value for each item according to the average ranking of the statement. Frequently the researcher will also calculate a Q-value for each item. This is determined by calculating an interquartile range of the distribution of judgments obtained for each statement. The interquartile range is a simple measure of ambiguity; it measures the spread of the middle 50 percent of the judgments on each single attitude statement. A high Q-value indicates that the judges do not agree on the placement of an individual item, and a low value indicates agreement. On the basis of scale value, the Q-value, and a desire to have items representing the full range of the eleven-point scale, the researcher chooses approximately 20 items that constitute the scale. The scale can then be given to research subjects and each subject's scale score is determined by the items with which he or she agrees. This final scale score is then treated as the measure of the respondent's attitudes and is used in the analysis.

Likert Scales

A second technique for developing scales was developed by Rensis Likert (1932). Likert's procedure, generally referred to as the **method of summated ratings**, is assumed to have ordinal rather than interval properties. Scores on Likert scales can be analyzed using any of the ordinal measures discussed earlier.

Likert's method also involves a series of steps. The procedure is begun in much the same fashion as the Thurstone procedure: the researcher assembles a large number of items relevant to the attitude content area. In this case, however, the items are written in such a way that degree of agreement or disagreement can be registered. Typically, a five-point scale is used with the following response categories: (1) strongly agree, (2) agree, (3) neutral or undecided, (4) disagree, and (5) strongly disagree.

The group of items is then presented to respondents who have characteristics that are similar to the population one is interested in studying. Each respondent is given a score, which is obtained by adding all of the responses together. For example, if a respondent strongly agrees with a given item, he or she is given a score of one, an agree response is given a score of two, a neutral response a score of three, and so on. Respondents

are then divided into quartiles from high to low and each individual attitude item is assessed in terms of whether or not it is able to distinguish between high and low scorers on the total set. A final set of about twenty statements that best discriminate between high and low scorers is then chosen to constitute the scale. These items can then be presented to the group that one is interested in studying.

For analysis purposes, scores on all of the items are generally added together and this sum constitutes the attitude score for each individual respondent. Scores can be ranked and ordinal statistics used for analysis purposes.

Guttman Scales

Guttman scaling was developed by Louis Guttman and his associates in their World War II studies of the American soldier. This technique, sometimes referred to as **cumulative scaling** or *scalogram analysis*, seeks to develop a set of items for attitude measurement that will be unidimensional.

Guttman sought to develop scales in which responses to any single item could be determined by a total score on the set of items. In other words, if a scale has Guttman properties, "an individual with a higher rank (or score) than another individual on the same set of statements must also rank just as high or higher on every statement in the set as the other individual" (Edwards, 1957: 172).

The property of **unidimensionality** can perhaps best be illustrated by Guttman scales of attitudes toward premarital sexual behavior. Reiss (1967) conceptualized premarital sexual behavior on a unidimensional continuum ranging from kissing to full sexual intercourse. The idea is that anyone who approves of heavy petting also approves of light petting and kissing, or any person who disapproves of light petting will also disapprove of heavy petting, oral contact, extra heavy petting, and full sexual intercourse. Total scores can be calculated for individual respondents, allowing the researcher to determine which individual items in the set the respondent agrees or disagrees with. Again, these scores are treated as having ordinal properties and can be analyzed accordingly.

There are many other scaling procedures used today in the social sciences. An important technique we will not describe here is the use of **factor analysis** to reduce large sets of indicators to one or a few composite scales. The important thing to remember, however, is that each technique is an exercise in data reduction; each is designed to combine responses to multiple items into a single score. That score then constitutes one's operationalization of the variable of interest. It is the scale score that is used in the analysis rather than the individual responses to single items.

As was indicated at the beginning of this chapter, whether one is

using single-item indicators or sets of items grouped into indices or scales, the final critical stage of the research process is interpretation. It is frequently at this stage that the skills of the researcher are most severely tested. Sense must be made of the research findings and meaningful conclusions must be reached. The findings and conclusions must then be presented in a meaningful research report. We now turn to the preparation of research reports.

VIII. SUMMARY

Once data relevant to some problem are collected, the next critical task becomes what to do with the data. The process of making sense out of masses of data is called analysis. In data analysis the researcher asks "What have we found?" Analysis typically follows several specific steps: (1) coding, in which verbal responses, written answers or accumulated records are converted to numbers; (2) data entry, in which coded data are punched onto computer cards or entered directly onto computer files; (3) descriptive analysis, in which the researcher examines frequency distributions for individual variables; (4) cross-tabulation, in which relationships between variables are examined; and (5) index construction, scaling, and multivariate analysis, in which more complex relationships among variables are treated.

Coding is the procedure by which the researcher converts responses to some standard form in preparation for analysis. The usual procedure is to develop a codebook which details the translation procedure. The codebook describes the specific steps by which a response to a question on a questionnaire, for example, is translated into a number that can be counted and compared with other numbers created through similar procedures. The reason for creating detailed codebooks is to make the coding process as mechanical as possible, so that researchers can easily translate responses into numbers or can enter numbers directly into computers. Coding procedures should be designed to retain a maximum amount of data but should also facilitate the analysis and interpretation procedures.

Data entry has become a rather straightforward step in data analysis. Coded numbers are entered onto computer cards or, more frequently now, directly onto magnetic cards or disks. Once entered and checked for errors of entry or coding, the data are ready for analysis.

The first step in analysis is usually descriptive. This involves a preliminary examination of the data to determine the distribution of responses on individual variables. For example, how many persons in a sample are married, how many are divorced, how many are plumbers or school teachers, and so on. Descriptive analysis might also involve computing percentages or calculating means, medians, or modes for the variables, and, often these distributions are compared with other distributions for that

variable from another time period or from respondents in other samples of the same or different geographic areas. The type and complexity of descriptive analysis frequently depend upon whether one's data are measured in nominal, ordinal, or interval scales. Different analysis procedures are applicable to each of these levels of measurement.

Cross-tabulations or contingency tables are created to compare the distributions of two or more variables simultaneously. The researcher asks: how is the distribution of responses on a dependent variable affected by the distribution of responses on the independent variable? Or: how are two independent variables interrelated? Creating cross-tabulations is a first step in assessing the type of relationship that exists among variables.

In addition to presenting one's findings in contingency tables, the researcher will usually want to report additional information such as whether an observed relationship is statistically significant or how much of the observed variation in a dependent variable is attributable to variation in one or more independent variables. To answer such questions, the researcher will use a variety of tests of significance or measures of association. The former answer the question of whether a relationship is "real" or could have occurred by chance; the latter specify the amount of variation in dependent variables that is explained by independent variables. Many different tests of significance and measures of association have been devised, and courses in statistics deal with the choice of appropriate tests or measures, their use, and their interpretation. Researchers select from among the available statistical tests according to what they want to do and the level of measurement they have achieved in their data collection.

Finally, part of data analysis usually involves data reduction. It becomes impossible for researchers to treat each indicator as a separate variable, and so items sharing some characteristic are combined into indices or scales. A single score representing many items is then used as a single variable. This procedure is widely used in the development of attitude scales, the most common of which are Thurstone, Likert, and Guttman scales. Factor analysis is also a common method of data reduction.

REFERENCES

- BLALOCK, HUBERT M., 1960. *Social Statistics*. New York: McGraw-Hill.
- CAPLOVITZ, DAVID, and FRED SHERROW, 1977. *The Religious Drop-Outs: Apostasy Among College Students*. Beverly Hills, Cal.: Sage.
- CAPLOW, THEODORE, and HOWARD M. BAHR, 1979. "Half a Century of Change in Adolescent Attitudes: Replication of a Middletown Survey by the Lynds," *Public Opinion Quarterly* 43 (Winter): 1-17.
- COSTNER, HERBERT L., 1965. "Criteria for Measures of Association." *American Sociological Review* 30: 341-353.

- LIKERT, RENSIS, 1932. *A Technique for Measurement of Attitudes*. Archives of Psychology, No. 140.
- LYND, ROBERT S., and HELEN M. LYND, 1929. *Middletown: A Study in American Culture*. New York: Harcourt & Brace.
- MUELLER, JOHN H., KARL F. SCHUESSLER, and HERBERT L. COSTNER, 1970. *Statistical Reasoning in Sociology*. Boston: Houghton Mifflin.
- REISS, I. L., 1967. *The Social Context of Premarital Sexual Permissiveness*. New York: Holt, Rinehart & Winston.
- THURSTONE, L. L., 1928. "Attitudes Can Be Measured." *American Journal of Sociology* 33: 529-554.
- U.S. Bureau of the Census, 1971. *1970 Census of Population, Alphabetical Index of Industries and Occupations*. Washington, D.C.: U.S. Government Printing Office.

Writing About Research

fourteen

I. Introduction

II. The Information Explosion

III. In Defense of Social Science Writing

IV. Writing as Scientific Creativity

V. Three Basic Issues in Writing

Audience identification

Organization

Style

VI. The Process of Writing

VII. Summary

I. INTRODUCTION

Presenting the findings is an essential stage of a research project. Whatever form the presentation takes—written report, a speech augmented by colored slides, executive summary and verbal responses to questions—clear, concise writing is the most critical component. As we have noted, good writing is also essential throughout a project: researchers must take notes on what they do, hear, or see; they must be able to describe things and transmit instructions or observations; and they must convince others that their work is significant. To paraphrase a recent guide to technical writing, researchers *must* write if they are to succeed (Barrass, 1978).

The body of this chapter has five parts. It begins with the point that scientists and almost everyone else profit from knowing the principles of good writing. The current information explosion heightens the need for effective communication because it increases the sheer bulk of writing that one should know about. Second, we discuss some “laws” about the ratio of quality work to dross, with special attention to the widely shared belief that the prose of social scientists is even worse than that of other scientists. Third, we treat writing as part of the *process* of scientific creativity rather than merely a means to pass on findings after the creative act. Fourth, we highlight three essential issues in scientific writing: audience, organization, and style. Finally, we depict the process of writing research and emphasize

that the road to a polished final product winds through numerous revisions. We conclude with a set of working principles to guide revision.

II. THE INFORMATION EXPLOSION

Researchers may complete a project, achieve interesting or even useful results, and yet fail because they do not or cannot communicate the findings to an appropriate audience. Unfortunately, some discoveries of great import remain unknown because they did not reach audiences that could understand and use them.

How important is writing about research? A chemist writes:

I was shocked to find that the time and effort of writing was often equal to that of the research work being described. In addition to research papers, I am continually involved in writing proposals, reports, course syllabi, laboratory experiment instructions, letters of recommendation, manuscript reviews, book reviews and books. This is the common experience of academic scientists and the tasks of writing fall heavily on industrial scientists as well. (Enke, 1978: 40)

Scientists often have trouble communicating with each other as well as with the public. Many universities now offer courses in technical or scientific writing, because while not everyone who studies research methods will write reports, almost everyone who follows a professional or semiprofessional career will have to write *something*—letters, memoranda, summaries of trips or meetings, evaluation reports, institutional materials.

Scientific writing in general, and the writing of social scientists in particular, has often been criticized as repetitive, unreadable, and needlessly obscure or esoteric. In a 1942 article entitled "The Sad Estate of Scientific Publication" T. S. Harding, an experienced editor, complained of the bad editing of scientific articles, the unfortunate fadism whereby "one outstanding paper by a major scientist breeds dozens of little parasitic papers, all equipped with cryptic mathematical formulas to fool the elect." He pleaded for a reduction in the rate of publication and for new methods to help individual researchers keep up with the burgeoning literature in their specialties. He concluded that unless something was done "to cope with the present disorderly flood of scientific publication," science and even Western civilization was threatened; the unmanageable flood would "simply drown science out" (Harding, 1942: 597, 601).

Professor Harding did not recognize the simplicity of his time. Figures on the growth of scientific literature in the 1940–1980 period reveal enormous increases in the number of scientific articles, journals, and books. In 1940 there were approximately 1,500 scientific and technical journals in the United States. In 1974 the *Science Citation Index* maintained by the Institute

for Scientific Information covered an estimated 401,000 articles from 2,443 scientific and technical journals (Garfield, 1980: 132). In 1977 the 2.8 million scientists and engineers in the United States produced 15,000 books, 4,500 journals and 4,500 other periodicals, as well as thousands of technical reports. Worldwide, 300,000 scientific and technical articles were published in 1977 (King, McDonald, and Roderer, 1981: 1, 9, 22).

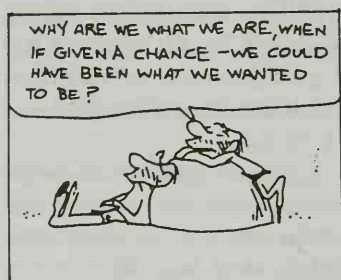
Fortunately, accompanying the increase in output have come more efficient methods for literature reviews, citation searches, and finding relevant work by scientists in disciplines other than one's own. Electronic data storage and retrieval systems have helped us deal with the "paper explosion" (Boutry, 1970) in the sense that literature reviews can be conducted far more efficiently and relevant work can be readily located. Among the computerized services offered by the world's most successful information-management company, the Institute for Scientific Information, are five services relevant to the world's journal literature in science and technology: (1) a weekly summary of tables of contents in six disciplinary areas, including a separate issue for social and behavioral sciences; (2) selective dissemination services which provide subscribers with weekly computer reports listing recent articles relevant to the subscriber's interests; (3) published volumes facilitating literature searches (*Science Citation Index* and *Social Science Citation Index*); (4) the availability of reprints or tearsheets from all articles covered by ISI; and (5) the periodical *Journal Citation Reports*, which permits librarians and others to identify "core" journals as revealed in counts of how often they are cited in the scientific literature (Garfield, Koenig, and DiRenzo, 1980: 288-290).

The advances in electronic archiving and literature searching have not been matched by advances in the quality of scientific writing. Today's scientific papers are not more readable than those of a generation ago. Most scientists have little training in literary composition. Furthermore, the pressures to make visible contributions to one's field, and thereby enhance the prestige of one's university or company or increase one's employment prospects, encourage the publication of material of low scientific merit and poor literary quality. The problem is widely recognized. One organizational attempt to deal with it are the many university courses in technical writing aimed at improving the writing skills of future engineers, chemists, and physicians.

III. IN DEFENSE OF SOCIAL SCIENCE WRITING

The social sciences have been singled out as leading offenders in the production of unreadable, complex prose. In part, this criticism reflects the subject matter of social science. One expects chemists and engineers to use terms not understood by the layperson. That is what their special education

ILLUSTRATION 14.1 An illustration of obscure communication.



B. C. reprinted with permission of Johnny Hart and Field Enterprises, Inc. Taken from Patricia Wright's "Five Skills Technical Writers Need," *IEEE Transactions on Professional Communication*, Vol. PC-24, No. 1, March 1981, p. 13.

was about—to allow them to understand mysteries not readily accessible to the ordinary person. The social scientist, on the other hand, studies ordinary things like churches, crime, family, and fertility.

When this ordinary stuff of life is relabeled in attempts to increase scientific clarity and facilitate classification, the ordinary reader may feel that the social scientist has renamed ordinary things to appear "scientific" about things practically everyone already knows.

Social scientists have sometimes used complex words when simple ones would do. Also, they write at least as badly as scientists in other disciplines. However, because the focus of social research is the everyday social world, social scientists are more likely to be seen as restaters of the

obvious or complexifiers of the mundane than are nuclear physicists, mathematicians, or philosophers, whose writing the general public does not expect to understand in the first place. Physical or biological scientists may produce dull articles that duplicate or slightly extend existing work and fool everyone but their more discriminating colleagues, because only they are expected to understand and evaluate the quality of research in physics or biology. However, people generally think they have a working knowledge about human society. Accordingly, when a critic translates a complex hypothesis on, say, the correlates of stratification into lay language that reads, "rich people have it better than poor people," the social scientist stands exposed as one who translates common clichés into fancy language.

Yet in social science, as in other disciplines, writers of grace and clarity are not hard to find. As models of precision and elegance in the use of language, several sociological classics come readily to mind: Robert and Helen Lynd's *Middletown* (1929) and *Middletown in Transition* (1937); Kenneth Boulding's *The Image* (1956); C. Wright Mills' *The Sociological Imagination* (1959); Peter Berger's *Invitation to Sociology* (1963); Elliot Liebow's *Tally's Corner* (1967); and Theodore Caplow's *Toward Social Hope* (1975), to name a few. To say that most scientists in most disciplines need to learn to write better should not obscure the point that every discipline, including the social sciences, has its share of writers of wit and style.

An especially biting and, regrettably, partially accurate critique of American social science is Stanislaw Andreski's book, *Social Science as Sorcery* (1972). He admits that social science has made numerous positive contributions to knowledge, but his book is explicitly devoted to the noncontributions, summarized in the position that "much of what passes as scientific study of human behavior boils down to an equivalent of sorcery" (Andreski, 1972: 10). Some of Andreski's chapter titles reveal the kind of sorcery he sees: Manipulation Through Description, The Smoke Screen of Jargon, Evasion in the Guise of Objectivity, Quantification as Camouflage, Ideology Underneath Terminology, and The Barbarian Assault on the Corrupted Citadels of Learning.

Our present concern is scientific writing, and the above sampling of chapter titles reveals that much of the "sorcery" stems from the misuse or illegitimate use of language. Aspects of communication—manipulative description, jargonic smoke screens, camouflagic quantification, and objectivity as disguise—are among the processes used by pseudo-scientific sorcerers and their apprentices to gull the naive public.

Among the critics who may be judged as helpful rather than caustic are writers like Samuel T. Williamson (1967), whose "How to Write Like a Social Scientist" includes a nice list of things to avoid in scientific writing. Despite the pejorative title, Williams claims his essay should not be seen as picking on social scientists, for, he says, their writing is no worse than that of most of their colleagues in other fields. His set of rules is not a bad

collection, really, when inverted and applied to improve one's writing. The basic rules are:

1. Never use a short word when you can think of a long one.
2. Never use one word when you can use two or more.
3. Put one-syllable thought into polysyllabic terms.
4. Put the obvious in terms of the unintelligible.
5. Announce what you are going to say before you say it.
6. Defend your style as "scientific." (Williamson, 1967: 112–113)

Similar satires on social science writing are common. Remus (1977) argues that complaints about the unreadability of scientific journals stem from readers' misunderstanding of what the journals are meant to do:

The readers think that the journals are written for them, but the journals are really for academicians to make points for promotion and tenure. Further, the reputation of a journal is a direct function of the obscurity of its pages, not its usefulness. (Remus, 1977: 65)

The ready availability of examples of mediocre writing can be turned to psychological advantage by the researcher willing to work at writing well. The novelist Stephen King told an interviewer that aspiring writers who did not read a lot were "in bad trouble." Asked why reading was so important, he replied:

The most important thing is it teaches you what *not* to do. I think young writers have reached a real watershed moment in their own lives as writers when they can say very honestly to themselves—maybe not even say it aloud—"I do better stuff than that." They read a book that's been printed and presumably somebody got paid money for it. Or if they didn't get paid money for it, they got copies to give to their relatives and friends. But when they read it, they make that vital critical judgment and say, "I'm better." (Janeczko, 1980: 10)

If the critics of contemporary social science research are right, the quality of craftsmanship necessary for a writer to able to say honestly, "I'm better" is readily attainable to one of even modest talent who seriously practices the craft of scientific writing.

Another advantage of the widespread perception that scientists generally write poorly is the growing availability of formal courses and workshops on technical writing, and a sizable literature on how to conduct such courses (Booth, 1979–1980; Gubanich, 1977). The flavor of the movement is illustrated in pieces entitled "A Mandatory Course in Scientific Writing for Undergraduate Medical Students" (Roland and Cox, 1976) and "Teaching Scientific Writing Humanistically: From Theory to Action" (Carlisle, 1978).

The guidebooks on how to write scientific reports range from stimulating collections of exemplary communication by great scientists who were also great writers (e.g., John Tyndall, T. H. Huxley, Charles Darwin, Isaac Newton, Benjamin Franklin, and Julian Huxley; see Ryan, 1960) to manuals on writing that readily acknowledge that in their "how to" approach they resemble cookbooks, the assumption being that anyone can produce an acceptable (edifying if not edible) product by carefully following the prescribed procedures in proper sequence (Day, 1979; Zinsser, 1976; Barrass, 1978). An example of the "cookbook" approach is Day's (1979) *How to Write and Publish a Scientific Paper*, which includes 26 brief chapters and six appendices on such topics as What Is a Scientific Paper? How to Prepare the Title, How to Write the Introduction, How to Cite the Acknowledgments, How to Type the Manuscript, How to Order and Use Reprints, Avoiding Jargon, and Common Errors in Style and in Spelling. The beginner will find such manuals useful. More experienced writers may prefer to spend their time analyzing exemplary pieces by talented writers and referring to more definitive guides for style and usage.

IV. WRITING AS SCIENTIFIC CREATIVITY

A piece of prose, be it literary or scientific, has a certain life of its own. As the writer encounters it again, while editing or revising, a kind of dialogue takes place that sometimes produces a creative restructuring. The sociologist Robin Williams has described how, in the course of writing a book, there came a point at which the book "had acquired a stubborn independence from the author's intentions" (Williams, 1976: 95). He tells that the emerging book began to "fight back." The interaction between craftsman and material is described as part of the poetic process, as a poet confronts a version of his or her past represented in a poem and interacts with that previous "vision" to produce a more satisfactory statement. The ensuing process of active craftsmanship is hardly different in science than in poetry:

It is helpful to observe what the poet works away from, what he avoids. When the poet deletes certain passages and strikes out single words or shifts pauses, he is not only demonstrating his mastery over the craft, but at times showing his impatience or dissatisfaction with a part of himself when he was younger, even if the vision under the knife of correction is only five minutes old. (Wallenstein, 1971: 7)

Albert Einstein said that "The whole of science is nothing more than a refinement of every day thinking" (Einstein, 1950: 59). Part of that refinement occurs in the process of writing about research. According to one science writer, "writing style is the principle by which everyday thoughts

are refined" (Aaronson, 1977: 13). Many of us remember speech classes where we were told, "If you can't say it, you don't know it." The process of writing serves a similar function in revealing whether we really know something. "Students must be convinced that they do not *know* what they mean until they can *say* what they mean," says Aaronson (1977: 14), who suggests that all undergraduate science majors be required to take a course in scientific writing.

Writing is part of the process of *generating*, rather than merely communicating knowledge, because it consists of joining bits of information into relationships.

Composing consists of joining bits of information into relationships, many of which have never existed until the composer utters them. Simply by writing—that is, by composing information—you become aware of the connections you make, and you thereby know more than you knew before starting to write. . . . No matter what the subject, no matter how much you might already know about it, simply writing about that subject will cause you to gain a new awareness of how the fragments of information about that subject relate to each other. This awareness is new knowledge. (Van Nostrand, 1979: 178)

Thus writing is a process of discovery. Effective writing about research reveals unanticipated relationships or highlights implications that were hazily if at all perceived before the process of writing forced systematic organization of findings, perceptions, and impressions. In other words, the writing necessary for the conduct of research, including the proposals, technical memos, and interim reports, as well as a final report, contributes to the clarification and organization of observations, which in the end produce the findings.

For example, making an interview schedule forces greater conceptual clarity on the researcher and requires decisions about priorities and resource allocation. Suddenly what seemed a simple question may become multidimensional and may dictate the need for increased attention to clarity and boundaries of concepts previously assumed to be unambiguous. Or, during the preparation of a codebook, the researcher may find that respondents have provided a richer, more detailed set of responses than was anticipated. The preparation of a codebook that allows researchers to deal with the unexpected complexities is itself a creative act that affects all subsequent stages of the project.

Later, preliminary presentations or reports of pretests and pilot studies may provoke queries or criticisms that shape the remaining stages of a project, or the preparation of interim reports may reveal unforeseen gaps or opportunities. Having to organize into coherent prose what researchers have done, what is presently underway, and what the remaining stages of the project look like prevents wasted effort and helps the researcher to

identify flaws or oversights while it is still possible to do something about them.

Perhaps most unsettling to the researcher who sets out to describe techniques and findings is the awareness that after all, he or she really cannot document what has gone on or what the actual results to this stage are. Having to describe and organize in hard copy forces the sensitive researcher to face the knowledge gaps that remain and highlights aspects of procedure or analysis that were not planned or implemented properly.

V. THREE BASIC ISSUES IN WRITING

Whether one is preparing a research report, a novel, or a letter, at least three topics need explicit attention: intended audience, organization, and style. Decisions related to each of these topics apply also to visual or spoken presentations, but we shall consider them mainly as they apply to the written research report.

Audience Identification

This initial stage of research writing is essential because decisions about organization and style depend on the audience. If one is writing to professional colleagues via publication in a scientific journal, there usually are constraints in organization and style imposed by the journal's editorial practices.

Audience identification also determines mode of presentation. For example, an in-house report prepared by the research staff of a private firm may be a verbal presentation supplemented by visual aids, a few illustrative handouts, and an executive summary. In contrast, a book aimed at the educated nonprofessional reader may require a different order of presentation, one designed to grab the reader's attention and hold it. A book on the same topic aimed at other professionals might contain technical details about procedure and theory undesirable in writing for a wider audience.

An audience of professional researchers may be very interested in the technical aspects of a study. On the other hand, the posture of managers, administrators, or the general public toward projects relevant to their interests is likely to be: "Don't bother us with these technical details; tell us what you found. It's the results that count; what are they, and what do they really mean to me?"

Organization

The essential question in organizing a research report is the arrangement of topical sections, or deciding what goes where. Despite a few critics' (Williamson, 1967: 13; Sachs, 1972) arguments that the repetition of telling

the reader what one is going to say, then saying it, and then summarizing what one has said adds to the dullness of scientific writing, we believe that a good report, article, or research monograph includes an introduction which serves as a roadmap to the rest of the report. Like the successful advertisement for a popular motel chain, which advertises that travelers who stay at their inns get no surprises, the well-written research report contains no surprises. Instead, at the outset a brief outline is offered, alerting readers to the nature and arrangement of the specific sections that follow. Thus, having been warned about the way the report is organized, the reader is mentally prepared for the sequence that unfolds in the body of the report and may skip sections of little interest and pay close attention to others.

The second stage in the three-stage model for organizing reports is the body, which includes a description of the research procedures and the findings. This central section may contain many or few subdivisions, but it must describe the researchers' logic, activities, and the apparent consequences of that thought and action.

Finally, there is the summary section, in which the writer reminds the reader once more of the stages in the mental journey just completed, emphasizes the major findings, interprets their meaning, and if appropriate, makes recommendations for action.

Thus the successful report, however it is organized within this three-part superstructure, prepares the reader by providing a kind of roadmap at the outset, describes the research processes and findings in appropriate detail, and then reviews the "trip," pointing to its high points and their meaning.

If we compare a scientific paper to a play, the introduction (Act I) provides a summary of the "plot" and introduces a "cast of characters" (generally the variables to be considered). The body of the paper (Act II) explains why these "characters" are worthy of the readers' time, their history, the relationship among them, how they are manipulated in the present "drama," and the results of that manipulation. The summary (Act III) is an epilogue that reviews the status of the "characters" and their interrelationships as revealed in Act II and may foreshadow a future drama growing out of the present one.

A more detailed outline simply provides details on the components of the three main parts. Whether all the subdivisions shown below are necessary is one of the choices the writer must make.

- Section I. **Introduction:** What the study is about; includes general objective.
- Section II. **Procedures and Findings**
 - A. Significance: Why this study is worth doing.
 - B. Previous work: What has been done on this problem by prior investigators.

- C. Specific objectives: Precisely what problems will be attacked.
- D. Research methods: The approach taken in confronting the problem, sometimes including separate sections on:
 - 1. Measurement of variables
 - 2. Sampling procedures
 - 3. Data-collection processes
 - 4. Techniques of analysis
- E. Findings, or the result of the application of the research methods. This section may be subdivided into categories of findings corresponding to the specific objectives or to particular variables or sets of variables.

Section III. **Summary and Conclusion**

- A. Discussion: A treatment of what the findings seem to mean, how the results relate to the specific objectives, and under what qualifications they may be accepted as valid.
- B. Summary and interpretation: In some papers this is the simple recapping of the findings with reference to the critical points mentioned in the "roadmap" at the introduction. In others it is a systematic treatment of the implications of the findings.
- C. Recommendations: Many reports do not contain recommendations. Others limit their recommendations to a section of the summary where needs for future work are described.

Style

The simplest definition of style is a comparative one: you know style when you see it. A research report written with style communicates ideas accurately, concisely, and interestingly (Hudson, 1972: 37). Style, like grace, is aimed at a recipient. It is conscious design aimed to ease or please.

Style in scientific reporting is in some ways a question of good manners. It means taking trouble to give the reader all he needs to know for a critical understanding of the text in a way which will make his task as easy for him as possible. (Buzzard, 1972: 202)

Albert Einstein's description of scientific conceptualizing is a description of style in action. The way the scientist uses concepts, Einstein said, is really not very different from the way we use them in daily life; the basic difference is "... merely in the more precise definition of concepts and conclusions; more painstaking and systematic choice of experimental material; and greater logical economy" (Einstein, 1950: 98). It is the painstaking and systematic choice of the proper words or concepts that marks good style. According to Jonathan Swift, "Proper words in proper places, make the true definition of a style" (Swift, 1720), and a modern critic adds that in addition to the right words in the right places, scientific style must be transparent, in that "the reader sees through the words to the underlying phenomena and concepts. The best scientific writing is characterized by brevity, clarity, and precision" (Aaronson, 1977: 6).

The topic of writing style cuts across both scientific and humanistic disciplines. There are several standard, first-rate books that anyone serious about writing style should have at hand for reference. We find Strunk and White (1979) particularly concise and useful. It contains a treatment of 11 "elementary rules of usage," 11 more "elementary principles of composition," chapters on "a few matters of form" and "words and expressions commonly misused," and a section entitled "An Approach to Style," which consists of discussion of 21 "reminders," such as "work from a suitable design," "revise and rewrite," "do not explain too much," "avoid fancy words," and "prefer the standard to the offbeat." Moreover, the entire book occupies merely 85 pages and probably contains more wisdom about writing per ounce of paper than anything else available.

Another masterful work is Jacques Barzun's (1975) *Simple & Direct: A Rhetoric for Writers*, which includes case analysis (e.g., "Aristotle on Detective Fiction") and the following chapters: Diction, or Which Words to Use; Linking, or What to Put Next; Tone and Tune, or What Impression Will It Make?; Meaning, or What Do I Want to Say?; Composition, or How Does It Hang Together?; and Revision, or What Have I Actually Said?

Finally, two highly recommended recent books on style and usage are Copperud's (1980) *American Usage and Style: The Consensus*, and Williams's (1981) *Style: Ten Lessons in Clarity and Grace*.

Barzun (1975: 3) is probably correct when he says that no one can teach another person to write but that writers are helped by criticism and that the would-be writer, using a book, can sometimes teach himself. In a well-known essay on English prose and style George Orwell (1954: 177) states that he is not concerned with "literary" use of language, but rather "language as an instrument for expressing and not for concealing or preventing thought." A good prose style is one that clarifies and stimulates. Here is Orwell's (1954: 176) brief list of rules for conveying rather than blurring meaning:

1. Never use a metaphor, simile or other figure of speech which you are not used to seeing in print.
2. Never use a long word where a short one will do.
3. If it is possible to cut a word out, always cut it out.
4. Never use the passive when you can use the active.
5. Never use a foreign phrase, a scientific word or a jargon word if you can think of an everyday English equivalent.
6. Break any of these rules sooner than say anything outright barbarous.

VI. THE PROCESS OF WRITING

Some researchers dictate their first drafts; others enter the draft into a word processor that shows a section of text on a computer console and permits instant editing; others type rough drafts and some write longhand. Experts

differ on how one ought to compose. Barzun (1945: 48) objected to having children learn to write on typewriters, but Lanham (1979: 53–54) insists that “the typewriter is mightier than the pen” in ease of revising one’s output. Whether one composes at the typewriter or not, he argues that the first draft needs to be typed before it can be revised properly:

The typewriter distances your prose and does it quickly. By depersonalizing our priceless prose, a typescript shows it to us as seen through a stranger’s eyes. It tells us what it looks like, literally how it “shapes up.” No single bad writing habit is so powerful as the habit of typing an essay only when you are ready to turn it in. Correct the handwritten manuscript by all means, but then type a draft and revise that. A typed version makes everything clearer, especially problems of sentence shape, rhythm, and emphasis. . . . “How do I know what I think until I see what I write?” enshrines a profound truth, and the typed draft allows this process to work most efficiently. So if you don’t know how to type, you must learn. For anyone who wants to write, typing is not a frill. It is essential. (Lanham, 1979: 54)

Research on the relative efficiencies of dictating versus writing suggests that authors can increase their speed of production from 20 to 35 percent if they dictate (Gould and Boies, 1978: 504), but the research also suggests that documents needing many revisions may be more efficiently written than dictated. The ultimate quality of the product does not seem to vary much whether it is dictated or written.

Whether the first draft is written or dictated is less important than the stages that follow. We stress the plural, *stages*, because good writing depends upon many revisions. We do not say that revision by itself can turn a poor writer into a good one. However, whatever the fraction of inspiration in one’s writing, careful revision will highlight that inspiration rather than allowing it to remain buried in sociologese, bureaucratese, or officialese.

We have talked about how research reports are organized—the necessary decisions about what goes where, the use of a standard outline running from introduction through recommendation. Given a draft which roughly conforms to the outline, the writer faces the task of revision. In our experience, revision is best accomplished by successive readings of the draft, each time examining the text for **economy**, for **elegance**, and for logical **evolution** from stage to stage. Some writers focus on one or two aspects of style in a single reading; other try to “cover the waterfront” as they rework a paper.

Economy in writing has to do with saying what must be said concisely. Economy is achieved by **pruning**, a process that involves taking each word, phrase, clause, and sentence by itself and asking: Does this say what I want it to say? Could I say it in fewer or better words? Is this bit, be it word, phrase, or clause, really necessary to my discussion?

Elegance has to do with both brevity and accuracy. Here we refer to the dictionary meaning of elegance in its mathematical sense, that is, scientific precision, neatness, and simplicity, rather than in its more frequent usage to denote refined grace or tasteful richness of design. The primary question is: Does the word, phrase, or sentence portray *accurately* as well as concisely what I wish to say? The proper choice of terms and their syntax—their arrangement within sentences and paragraphs—are essential to elegant writing. The writer will use a thesaurus or a dictionary and will carefully weigh the connotations of possible synonyms. Other things being equal, he or she will choose the shortest word or phrase that conveys as precisely as possible the intended meaning. There is much attention to nuance and tone, to implication and assumption, as one edits for elegance.

Evolution as an ideal in writing has to do with flow, transition, and logical connection between one passage and another. It applies to the smooth unfolding or development of the piece, to the ease with which a reader is carried along without confronting any missing links. Editing with attention to the flow of the prose removes choppy, disconnected ideas or links them with smoothing transitions so that the reader is never brought up short, suddenly wondering how he or she got to this point and what it has to do with the material that preceded it.

Proficiency in pruning, whether in writing or in horticulture, comes with practice. Great writers may be born, not made, but *good* writing is in the reach of most researchers. The key to effective writing, say the experts, is practice: "The best way to learn is to write, and then write some more" (Hudson, 1972: 37).

Our own pruning is guided by a list of working principles built on ideas presented in this chapter, a set of guidelines we have found helpful. Other writers might delete some of these principles and substitute others. The principles all apply only in the context of "other things being equal"; that is, there may be good reasons for violating each of them. As general principles, however, they help move prose toward economy, elegance, and evolution.

1. Short is better than long. This applies to words, sentences, paragraphs, papers, chapters, books, and scientific communications of any sort.
2. Active voice is better than passive.
3. Old words are better than new ones.
4. Clarity is more important than literary variety; use the same word to refer to the same thing.
5. Simple is better than complex.
6. System is better than ambiguity or confusion: in practice, text divided into headings and subheadings, thereby highlighting organizational form, is better than text in which form is implicit.
7. Surprise is generally bad form; build expectations and meet them.

8. Tables, figures, or other illustrations worth including usually merit discussion in the text.
9. Tables, figures, and other illustrations should be simple enough to be interpreted and understood apart from the text.

VII. SUMMARY

Early in this chapter we said that social scientists were frequently singled out by critics as exemplars of obscure, jargon-filled prose that belabored common-sense notions about human society. Even distinguished authors of manuals on style and usage berate the poor sociologist.

It is true that improvements in the writing of scientists generally have not kept pace with the computer-aided improvements in the methods of indexing, archiving, and literature-searching. There are more scientists now living than ever before, and consequently the volume of mediocre papers and presentations is larger; but so, we believe, is the number of excellent pieces. We would like to be able to point to a positive trend in writing about research; to say that the proportion of mediocre papers being published was declining.

Perhaps great writing cannot be taught, but organized, concise writing which systematically treats issues in an orderly, comprehensible way *can* be. We can improve the quality of scientific writing by encouraging revision by recipe. Correct usage and, to some extent, correct style can be conveyed. We listed several of the good "recipe books" in this chapter, and the chapter itself aimed at producing some guidelines for writing.

If the poet, whose reliance on inspiration and fleeting insight is far more developed than that of the social scientist, is shown to be "builder, mechanic, and craftsman, as well as thinker, 'feeler,' and creator" (Wal-lenstein, 1971: 5), then perhaps we can hope to mold the social scientific muse by paying more attention to ourselves as craftsmen, builders, and mechanics in the trade of prose construction.

REFERENCES

- AARONSON, STEVE, 1977. "Style in Scientific Writing." *Current Contents* 9, No. 2: 6-15.
- ANDRESKI, STANISLAV, 1972. *Social Science as Sorcery*. London: Andre Deutsch.
- BARRASS, ROBERT, 1978. *Scientists Must Write*. New York: Halsted Press.
- BARZUN, JACQUES, 1945. *Teacher in America*. Boston: Little, Brown.
- , 1975. *Simple and Direct: A Rhetoric for Writers*. New York: Harper & Row.
- BERGER, PETER L., 1963. *Invitation to Sociology*. Garden City, N.Y.: Doubleday Anchor Books.

- BOOTH, VERNON, 1979/80. "Scientific Writing in English." *Journal of Research Communications Studies* 2: 107-111.
- BOULDING, KENNETH E., 1956. *The Image*. Ann Arbor: University of Michigan Press.
- BOUTRY, G. A., 1970. "Quantity versus Quality in Scientific Research (II): The Paper Explosion." *Impact of Science on Society* 20, No. 3: 195-206.
- BUZZARD, R. B., 1972. "Notes on Report Writing," *Occupational Psychology* 46: 201-207.
- CAPLOW, THEODORE, 1975. *Toward Social Hope*. New York: Basic Books.
- CARLISLE, E. FRED, 1978. "Teaching Scientific Writing Humanistically: From Theory to Action." *English Journal* 67 (April): 35-39.
- COPPERUD, ROY H., 1980. *American Usage and Style: The Consensus*. New York: Van Nostrand Reinhold.
- DAY, ROBERT A., 1979. *How to Write and Publish a Scientific Paper*. Philadelphia: ISI Press.
- EINSTEIN, ALBERT, 1950. *Out of My Later Years*. New York: Philosophical Library.
- ENKE, C. G., 1978. "Scientific Writing: One Scientist's Perspective." *English Journal* (April): 40-43.
- GARFIELD, EUGENE, 1980. *Essays of an Information Scientist*, Vol. 3, 1977-1978. Philadelphia: ISI Press.
- GARFIELD, EUGENE, M. KOENIG, and T. DIRENZO, 1980. "ISI Data-Base-Produced Information Services." Pp. 288-298 in Eugene Garfield, *Essays of an Information Scientist*, Vol. 3, 1977-78. Philadelphia: ISI Press.
- GOULD, JOHN D., and STEPHEN J. BOIES, 1978. "How Authors Think about Their Writing, Dictating, and Speaking." *Human Factors* 20, No. 4: 494-505.
- GUBANICH, ALAN A., 1977. "Writing the Scientific Paper in the Investigation Lab." *American Biology Teacher* 39 (January): 27-34.
- HARDING, T. SWANN, 1942. "The Sad Estate of Scientific Publication." *American Journal of Sociology* 47 (January): 593-601.
- HUDSON, GEORGE R., 1972. "Making Evaluation Reports Readable." *Social Counselor* (September): 37-41.
- JANECZKO, PAUL, 1980. "In Their Own Words: An Interview with Stephen King." *English Journal* 69, No. 2: 9-10.
- KING, DONALD W., DENNIS D. McDONALD, and NANCY K. RODERER, 1981. *Scientific Journals in the United States: Their Production, Use, and Economics*. Stroudsburg, Pa.: Hutchinson Ross.
- LANHAM, RICHARD A., 1979. *Revising Prose*. New York: Charlie Scribner's Sons.
- LIEBOW, ELLIOT, 1967. *Tally's Corner: A Study of Negro Streetcorner Men*. Boston: Little, Brown.
- LYND, ROBERT S., and HELEN M. LYND, 1929. *Middletown: A Study in American Culture*. New York: Harcourt, Brace.
- 1937. *Middletown in Transition: A Study in Cultural Conflicts*. New York: Harcourt, Brace.
- MILLS, C. WRIGHT, 1959. *The Sociological Imagination*. New York: Grove Press.
- ORWELL, GEORGE, 1954. *A Collection of Essays*. Garden City, N.Y.: Doubleday.
- REMUS, WILLIAM, 1977. "Strategies for a Publish or Perish World or Why Journals Are Unreadable." *Interfaces* 8 (November): 64-68.
- ROLAND, CHARLES G., and BARBARA G. COX, 1976. "A Mandatory Course in Scientific Writing for Undergraduate Medical Students." *Journal of Medical Education* 51 (February): 89-93.
- RYAN, LAWRENCE (ed.), 1960. *A Science Reader*. New York: Holt, Rinehart & Winston.
- SACHS, HARLEY L., 1972. "Seduction of the Reader's Mind: The Technical Article as Strip-Tease." *Technical Communication* (Second Quarter): 12-13.

- STRUNK, WILLIAM, JR., and E. B. WHITE, 1979. *The Elements of Style*, 3rd ed. New York: Macmillan.
- SWIFT, JONATHAN, 1720. "Letter to a Young Clergyman," 9 January. Quoted in J. Bartlett, *Familiar Quotations*. Boston: Little, Brown, 1968, p. 389.
- VAN NOSTRAND, A. D., 1979. "Writing and the Generation of Knowledge." *Social Education* (March): 178-180.
- WALLENSTEIN, BARRY, 1971. *Visions and Revisions: An Approach to Poetry*. New York: Crowell.
- WILLIAMS, JOSEPH M., 1981. *Style: Ten Lessons in Clarity and Grace*. Glenview, Ill.: Scott, Foresman.
- WILLIAMS, ROBIN M., JR., 1976. "A Neglected Form of Symbolic Interactionism in Sociological Work: Book Talks Back to Author." *American Sociologist* 11 (May): 94-103.
- WILLIAMSON, SAMUEL T., 1967. "How to Write Like a Social Scientist." Pp. 109-114 in Raymond J. Pflug (ed.), *The Ways of Language*. New York: Odyssey Press.
- ZINSSER, WILLIAM, 1976. *On Writing Well: An Informal Guide to Writing Nonfiction*. New York: Harper & Row.

The Utilization of Social Science Research

fifteen

I. Introduction

II. Some Uses of Social Science Research

III. Social Science Research and National Policy Making

Report of the National Advisory Commission on Civil Disorders

Violence in America: The National Commission on the Causes and Prevention of Violence

The equality of educational opportunities

Social science research and decision making

IV. Barriers to Social Research Utilization

Problems inherent in the research enterprise

The controversial nature of social research

Problems with prediction

Belaboring the obvious

V. Social Research and the Individual Consumer

*Some additional cautions
the methodological screen
the conceptual screen*

The business of science

VI. Summary

I. INTRODUCTION

We argued in Chapter 1 that this book was written for *producers* and *consumers* of research. Our general assumption was that few social science students taking their first methods course will ultimately become social science researchers—though we also noted that most people are applied social scientists to some degree. Moreover, virtually all students who take such a course will be consumers of research throughout their lives. We “consume” research when we read the evening paper, watch television newscasts and documentaries, make decisions about investments, or consider the relative advantages of consumer products.

Furthermore, very little that affects us in our everyday lives is unaffected by social research. What the political candidate chooses to tell, or not to tell, is carefully determined from an analysis of research data designed to inform the candidate about constituents’ preferences. The products available in our supermarkets often are not marketed until people’s reactions and preferences have been carefully gauged through consumer research. The list could go on and on. All of us are affected almost daily by social science research.

In this final chapter we treat the question of where and how social science research is used. We begin with the assumption that most research is undertaken to answer questions. The answers may solve some important

problems, improve conditions, assist in the marketing of products, improve the functioning of an organization or program, assist decision-makers, and so on. The important thing is that someone has a question to be answered and it is believed that information provided by research will lead to a better, more informed, more acceptable, or more profitable answer to the question than is available without research.

Most social scientists tend to feel that social science research has been underutilized, and most would probably agree that one of the problems they face is to convince potential users that the findings of social science research are useful. They feel that they have a "product" that can play a more important role in helping people make better decisions because they are better informed. Elbridge Sibley summarizes this view:

Throughout its history, sociology has been the butt of slings and arrows from both academicians and the lay public, and at the same time it has suffered—probably more than most disciplines—inner doubts of its identity. Nevertheless, most of its exponents have been moved by melioristic impulses and have clung tenaciously to their faith that an objective, scientific, and ethically neutral study of society could, directly or indirectly, contribute to the improvement of the human condition. (Sibley, 1971: 13)

This same recognition of the potential of social research has been echoed by others. For example, the Behavioral and Social Sciences Survey Committee appointed by the National Academy of Sciences and the Social Science Research Council discussed what they viewed as the important future role of the social sciences: "The social sciences will provide no easy solutions in the near future, but they are our best hope, in the long run, for understanding our problems in depth and for providing new means of lessening tensions and improving our common life" (quoted in Sibley, 1971: 14). Weiss has similarly argued:

From the emergence of the social sciences, each of them has had a lusty strand of usefulness and application. Macchiavelli's *The Prince*, Adam Smith's *Wealth of Nations*, and Auguste Comte's *Positive Polity* were deliberately oriented to the solution of practical, worldly problems. It would be a strange kind of social science indeed that did not take the world around it not only as learning ground but as practice ground for its developing insights. While every science goes through periods when primary attention goes to developing tools and methodologies, to building bodies of data, and to constructing models and theories, the strain toward service is never absent. The social sciences are particularly susceptible to calls to the cause of human betterment. (Weiss, 1977: 3)

On the other hand, social scientists already play a much larger role as consultants, advisors, and staff technicians than they once did. Horowitz (1971) argues that a research role for social scientists has been institutionalized as part of the policy-making process and the question is now less

one of the scientific status of the social sciences and more one of how social and political research is used and whose interests it ultimately serves.

In this chapter we will try to illustrate the diversity of possible applications of social science research knowledge. We will note areas where such research findings have been used, the problems that limit broader usage, and some of the major areas where such knowledge could be used, particularly by the individual consumer. We will consider how the individual, acting outside of any formal role in the research process, can benefit from the results of social science research.

II. SOME USES OF SOCIAL SCIENCE RESEARCH

Several years ago Lazarsfeld, Sewell, and Wilensky (1967) were commissioned by the American Sociological Association to write a book on the uses of sociology, especially in nontraditional, nonacademic settings. The product of this effort was a lengthy collection of essays that described the role sociologists and their work have played in medicine, social welfare, business management, the military, law enforcement, foreign policy, urban planning, international and Third World development, and so on. The list highlights ways that social scientists have penetrated modern society and now play a role that is much broader than that of basic researchers in academic communities.

Despite the range of organizations, agencies, and programs identified by Lazarsfeld, Sewell, and Wilensky, the specific roles played by the social scientist appear somewhat more narrow. Generally, these have involved:

1. *Sharing with an agency, program, or organization the general wisdom or knowledge that the researcher has gleaned from social science research.* For example, social scientists might discuss with management of a large organization what has been learned about the differential effectiveness of alternative forms of management structure.
2. *Doing a special study.* This can involve a broad range of activities such as completing an attitude survey for a local Chamber of Commerce, conducting an experiment on the effectiveness of alternative incentive systems, or evaluating the impacts of a new educational program.
3. *Sensitizing personnel to sociological orientations.* In this role the sociologist has been involved in such activities as training managers to be more aware of the social needs and preferences of their clients, helping program administrators to recognize social factors that can affect outcomes of their programs, and so on.

Another view of alternative uses can be found in the work of Weiss and Bucuvalas:

People can use social science research to clarify the relative advantages of alternative choices, but they also use it conceptually; to understand the background and context of program operation, stimulate review of policy, focus attention on neglected issues, provide new understanding of the causes of social problems, clarify their own thinking, reorder priorities, make sense of what they have been doing, offer ideas for future directions, reduce uncertainties, create new uncertainties and provoke rethinking of taken-for-granted assumptions, justify actions, support positions, persuade others, and provide a sense of how the world works. (Weiss and Bucuvalas, 1980: 305)

In addition to the work being done in such important areas as law and medicine, social scientists have become involved in the development and implementation of public policy. In such settings, social scientists may be involved in the actual design of programs rather than in the analysis of their operation. Some of the major problems faced by the social science researcher in the policy-making arena are detailed in an excellent series of papers compiled by Horowitz (1971). These will receive additional attention when we examine some of the barriers to the use of social science research. Certainly social scientists are playing a broader role in policy development, program evaluation, and social impact analysis than they did in the past.

III. SOCIAL SCIENCE RESEARCH AND NATIONAL POLICY MAKING

Compared with the experience of some of the physical and biological sciences, there are relatively few instances of the application of social science research findings in national policy making. This does not mean that social science research has not played an important role in affecting some public policy and, as we shall see, its influence seems to be increasing, partly because of the now mandated role of social science research in such areas as social impact assessment.

During the late 1960s, social science research received extensive attention in the form of two National Commission reports, the first dealing with civil disorder and riots in the cities and the second on the related issue of violence in America. It will be remembered that this was a particularly difficult time in American history. Hundreds of cities had witnessed racial violence during the summers following the Watts riot of 1965, personal violence against public leaders had reached a peak in the assassination of Senator Robert Kennedy and civil rights leader Martin Luther King, and millions of Americans were fearful for their own personal safety in view of spiraling crime rates. People were looking for explanations and solutions, and they sought help from the social sciences. The contributions of the social science community are summarized in two highly publicized reports, *The Report of the National Advisory Commission on Civil Disorders* (1968) and

Violence in America (1969), a report of the Commission on the Causes and Prevention of Violence.

Report of the National Advisory Commission on Civil Disorders

The summer of 1967 brought extensive racial disorder and violence to many American cities, a continuation of the pattern of the previous two summers, following the August 1965 riot in Watts. In the summer of 1967 almost 150 American cities reported disorders of some magnitude in minority (usually black, though sometimes Puerto Rican) neighborhoods. The worst came in a two-week period in July when there were major riots in Newark and in Detroit. In reaction, President Lyndon Johnson appointed a National Advisory Commission on Civil Disorders which was charged to answer three basic questions: (1) what happened, (2) why did it happen, and (3) what could be done to prevent it from happening again? The Commission's summary report, generally referred to as the Kerner Commission Report after Illinois Governor Otto Kerner who served as chair, is some 600 pages long and includes extensive summaries of social science research on race and ethnic relations, prejudice and discrimination.

The Kerner Commission concluded that the nation was rapidly moving toward two increasingly separate Americas and that, without action, this division could become too deep to bridge. It further argued that such a division would lead to sustained violence in the cities as well as the danger of conclusive repudiation of "the traditional American ideals of individual dignity, freedom, and equality of opportunity." The report recommended action on a national scale that would include the immediate creation of new jobs and job training opportunities for the poor and the ethnic minorities, increased efforts to eliminate *de facto* racial segregation through substantial federal aid to school systems, reform of the welfare system that would guarantee standards of assistance at least as high as the annual "poverty level" set by the Social Security system, and enactment of a comprehensive, enforceable federal open housing law.

Each conclusion and recommendation in the report was based, at least in part, on social science research. Unfortunately, relatively few of the commission recommendations were ultimately translated into policy by the Johnson administration. We shall note some of the important reasons for this when we review the major barriers to the use of social science research.

Violence in America: The National Commission on the Causes and Prevention of Violence

Civil disorders and urban riots were only part of the pattern of increased concern about violence in the late 1960s, both at the individual and the group level. In addition to the work of the Kerner Commission, Dr.

Milton Eisenhower was chosen to head a second commission charged to "go as far as man's knowledge takes" it in searching for both the causes of violence and the means of preventing future violence. One strategy of this commission was to assemble a large advisory committee of social scientists. One of the most important products of this effort was the compilation of a series of reports which provide a comprehensive overview of the history and causes of violence in American society. The individual reports examined such questions as the role of the frontier and vigilante traditions as antecedents of current patterns of American violence, as well as the history of labor violence, racial violence, and war.

The primary conclusion of this series of reports was that violence was not new, it had only been "rediscovered." One author summarized: "Historically, collective violence has flowed regularly out of the central, political processes of western countries. Men seeking to seize, hold, or realign the levers of power have continually engaged in collective violence as part of their struggles. The oppressed have struck in the name of justice, the privileged in the name of order, those in between in the name of fear" (Tilly, 1969). While the report presented an impressive summary of research on violence, it concluded that comparatively little was known about how civil *peace* is created and maintained. Yet the latter is the very type of social science knowledge that is most needed and that, we assume, could be most readily used.

The Equality of Educational Opportunities

Perhaps the most extensive use of social science research has been in the areas of race relations and education. Among the most widely cited and controversial studies in these areas is a report dealing with the equality of educational opportunity authored by James S. Coleman and several colleagues that was prepared for the United States Department of Health, Education, and Welfare (1966). It examined the quality of educational opportunities available in the public schools to persons of different racial backgrounds. The report stated that the school one attends affects educational achievement primarily through the quality of the teachers and through the educational aspirations and backgrounds of the students attending the school.

The most disturbing finding of the study was the underachieving of minority students as demonstrated on standardized achievement tests. The report concluded that since the quality of schools was similar, then *segregation* must be the principal cause of minority underachievement. It was suggested that ethnic segregation forced minority students to attend schools with their ethnic peers who also came from deprived socioeconomic backgrounds. These peers have attitudes about school and themselves and have behavior problems that interfere with academic performance. The minority

classroom was characterized by apathy, despair, and failure. School integration was viewed as a way to disperse minority students among middle-class achieving students, thereby creating a better learning environment for them.

The report concluded that school integration should increase the educational achievement of black children. This finding became one of the bases of support of advocates of busing as a way to increase the educational opportunity of black students.

An unanticipated consequence of increased efforts toward school integration following the Coleman report was a so called "white flight" of families from cities to the suburbs to avoid sending their children to integrated schools. Coleman (1975), author of the original *Equality of Educational Opportunity* report, later analyzed enrollment records from the 20 largest central school districts for 1968–1973 and concluded that court-ordered desegregation, especially if busing was involved, had caused white families to flee. He argued that in the long run integration programs had created even greater segregation than had existed before the attempt. The reaction to Coleman, the former champion of school integration, was swift and vociferous. Critics (Green and Pettigrew, 1976) argued that Coleman had made serious methodological errors and that his conclusion was erroneous. Whether white flight had or had not occurred remains a point of debate. What is important is that the Equality of Educational Opportunity study had and continues to have major impacts on public education in American society.

Social Science Research and Decision Making

In at least one important arena we have recently entered a new era in the utilization of social science research. As was discussed in Chapter 12, the National Environmental Policy Act in 1969 mandated that a comprehensive and systematic evaluation be conducted in regard to major actions that would affect federal lands or other areas over which the federal government has regulatory or management responsibility, to determine the effects of such actions on both natural and human environments. Thus, interdisciplinary research that specifically includes the integration of natural and social sciences in planning and decision making is required by law.

In effect, this mandate created a supportive new environment for social science research. Previously social scientists had had to make their case for the need of social impact analysis on a project by project basis. Today, social research is an integral part of the social impact assessment process, and its findings affect decisions as diverse as whether and where the proposed MX missile system should be located, what the Department of Interior ought to do with the growing population of wild burros on

federal lands in the West, how many visitors the National Park Service should allow to travel by boat through the Grand Canyon, alternative strategies for the management of National Forests, and whether proposed nuclear power generating facilities should be licensed. Social science data are also widely used by state agencies in their licensing, permitting, and siting decisions for major projects such as electrical power plants and nuclear waste repositories, and by private industry both in permit application processes and in site selection for proposed projects.

The relative newness of work in this area makes a detailed evaluation of its importance and effectiveness difficult. However, there is no doubt that in social impact analysis a research role for social scientists has been institutionalized.

IV. BARRIERS TO SOCIAL RESEARCH UTILIZATION

We have identified several contexts where social science research contributes its findings and perspectives. Despite such successes, there are still some barriers that inhibit the use of social research findings. Among the most important of these barriers are (1) problems inherent in the research enterprise itself, (2) the controversial nature of some social science research findings, (3) the charge that social science provides less predictability or assurance of validity and reliability than do the "hard" sciences, and (4) the apparent obviousness or "common sense" nature of many of the findings of social research.

Problems Inherent in the Research Enterprise

Gans (1971) identifies an important barrier to the broader use of social science research which is directly a function of the scientific enterprise itself. He notes that policy researchers must be concerned with changing society rather than merely understanding it. Social scientists tend to be trained as basic researchers—that is, their focus is really on *understanding* society. Typically, they have felt that the task of changing society belongs to others. As a result, the development of proposals of "deliberate activity to affect the workings of society or any of its parts" are foreign to most social scientists.

The problem is compounded, Gans suggests, by four related factors that affect the credibility and relevance of social research. First, the self-concept of the academic social scientist often includes an objective detachment from society and such detachment is not very useful for policy-oriented research because "it generates theories and concepts more suitable for the bystander to the social process than the participant" (Gans, 1971: 20).

A second problem is a tendency by theorists and researchers to identify the broad impersonal causal forces that underlie most practical social problems. However, users of research often need ideas that are less universal and less impersonal. Gans notes that while urbanization and industrialization may very well have had major effects on the structure of the family, persons who must design family policy can do little with such abstract or general processes and need concrete variables and intervention points.

A closely related third problem is that social science theories tend to be so general that they are difficult to apply to specific problems or situations. For example, what specific aspects of urbanization and industrialization affect the structure of the family and in what ways?

Finally, the academic social scientist tends to operate at a level of abstraction that produces concepts difficult to apply to the real-life situations where the policy designers must work. Before social science research data can be applied, potential users have to consider a study worthy of attention and give it a hearing. Decision-makers generally use two basic tests to screen incoming research findings, a truth test and a utility test:

. . . the truth test is composed of two independent dimensions. Decision-makers appraise a study in terms of its technical merit, but they have another basis for screening its truth claims—its conformity with their prior understanding and experience. . . . The utility test, too, is made up of two analytically distinct components. One is the extent to which a study provides explicit and practical direction on matters they can do something about . . . [The other] encompasses a wide range of conceptual contributions, from helping to "establish new goals and bench marks of the attainable," to helping to "enrich and deepen understanding of the complexity of the problems and the unintended consequences of action," to guiding the "effort to interpret and structure the social world by establishing languages and symbolic universes in comprehending and carrying on social life." (Weiss and Bucuvalas, 1980: 308–309)

Much social science research may not be used by administrators because it fails to pass one or both of these important tests. A closely related problem is the rather "frail character" of the knowledge produced by social scientists. As Weiss has noted:

Government officials know that social science is beset with fads of attention, with competing theoretical frameworks, and with contradictory empirical evidence. When confronted with the latest words from the social science oracle, they retain a healthy skepticism. On their side, social scientists, too, are aware of the fragility of their knowledge. As a consequence, they often tend to be timorous about drawing policy implications from their work and reluctant to give clearcut recommendations. Because of the shaky edifice on which they are perched, they lack the zest and confidence of princes' counselors. (Weiss, 1977: 10)

A rather different set of problems has been identified by several academically oriented social scientists. These problems are typically referred to as "risks" that are assumed to constitute some threat to the academic community should that community increase its involvement in policy-making research. Illustration 15.1 summarizes the most frequently identified of these risks. These include concerns about government sponsorship determining the types of problems that social scientists will address, assessing those problems from the value orientation of the sponsors, and so on.

These four characteristics of the research enterprise act as barriers to more effective and extensive utilization of research findings. They also identify topics of concern to social scientists who desire to make more effective input in policy-related research.

The Controversial Nature of Social Research

We have already discussed such major projects as the Kerner Commission Report, the Violence Commission Report, and the Coleman Report. All of these reports were highly controversial, not only in the political arena but also in the scientific community. The fact that more of the various

ILLUSTRATION 15.1 RISKS TO THE ACADEMIC COMMUNITY SOMETIMES ASSOCIATED WITH POLICY RESEARCH

1. When government sponsors social science research for policy purposes, it diverts social scientists from their true priority of enlarging knowledge.
2. When social scientists accept government funds, they are put in a position of giving advice prematurely on the basis of inadequate knowledge. This situation benefits neither government nor the social sciences; in fact, their brashness brings the discipline into disrepute.
3. By accepting research problems set forth by government officials, social scientists distort the development of the disciplines. Those areas in which government is interested are advanced, while issues of greater salience for the theoretical core of the discipline are neglected.
4. Social scientists who not only address the research problems that officials set but who address them in the terms and with the value orientation of the incumbent officeholders abdicate their role as scientists and become technicians for the powers-that-be.
5. Using their knowledge and skills in service to any sitting government—whether it be conservative or liberal and reformist or socialist—makes social scientists handmaidens of the state, rather than what they should be: unattached, untrammelled critics of any and all forms of social arrangement.

Source: Weiss, *Using Social Research in Public Policy Making*, 1977, p. 2, Lexington, Mass.: Lexington Books.

committees' recommendations were not translated into policy is attributable to the fact that most of the recommendations were contrary to the interests of powerful segments of the population. In each case, special interest groups found much in the reports that was disagreeable or threatening.

The findings of social research are often controversial or threatening. Defense contractors have often resented being objects of study, partly because corruption, waste, and mismanagement have often been uncovered in such investigations. Similarly, police departments have sometimes hesitated to open their records to social scientists because prior studies have documented racist practices or other inequities in police procedure.

In fact, social research often challenges an existing status quo and is thus threatening to persons who benefit from that status quo. Accordingly, even when social research is permitted, the findings are often ignored or shelved. Nevertheless, Peter Berger argues that sociology (and, by extension, the social sciences) are "justified by the belief that it is better to be conscious than unconscious and that consciousness is a condition of freedom" (Berger, 1963: 175).

Problems with Prediction

A most common criticism of the social sciences is that their findings have little predictive validity. A mature science should be able to offer accurate predictions about the relationships among the variables within its purview. However, prediction adequacy is evaluated rather differently in the physical than the social sciences. Henshel explains:

Physical science is commonly employed to predict the reactions of *systems that have been constructed according to its specifications*: social science is expected to predict "undisturbed" phenomena and to do so in areas that are either beyond human control or are controlled by nonscientific personnel (for instance, to produce "reformed" criminals in an institutional framework created by nonscientists, with aims other than reformation in mind, and run by nonscientific personnel, again with aims other than reformation). . . . Physical science, in short [and in contrast], has won its great prestige in prediction on systems it has itself created. (Henshel, 1971: 215)

Henshel argues that when one compares the accuracy of predictions by physical scientists in natural or uncontrolled physical systems, the results are not much better than those achieved by social scientists. The important difference between constructed and natural systems is the amount of control the researchers can maintain over the variables under study. Uncontrolled (and frequently uncontrollable) confounding variables in the natural system are eliminated, where possible, in the constructed system.

A frequent criticism of the social scientist who tries to use the methods of natural science and constructs the experimental situation is that one

cannot duplicate the complexities of social phenomena in ways that the findings have external validity, that is, that they will be relevant outside the artificial setting of the laboratory. However, Zelditch (n.d.) and Carlsmith, Ellsworth, and Aronson (1976), among others, have argued that objections of artificiality directed against social science research in laboratory contexts are based on a mistaken view of the purpose of such studies. The purpose of experiments, as well as of other types of research, is to study the critical aspects of a problem. Generalization is carried out on the basis of theory rather than on the extent to which one replicates some aspect of the "real world" in the laboratory. Whether artificiality or realism hinders research depends largely on the experimenter's goal. Weick (1965: 207) suggests: "Clearly, there is little warrant to the assumption that the greater the similarity between the experiment and the field situation, the greater the generality of the experimental outcomes."

We certainly do not argue for the establishment of a "constructed" social system as a means of testing the predictability of sociological theories of human behavior. However, it may be argued that if a highly controlled experimental setting is used, the "predictability" record of the social scientist is not very different from that of the natural scientists. The question often becomes one of how much emphasis should be placed on one or the other—predictability or control. Henshel (1971: 216) says that "lack of consensus on social goals portends a continued scarcity of sociologically directed constructed systems and hence an unfavorable prognosis for sociology's predictive prowess in the near future." Constructed systems created in the minds of scientists and philosophers (i.e., *Brave New World* and *Walden II*) exhibit substantial predictability, but at present, human societies seem willing to sacrifice predictability at a high level for a healthier dose of "freedom and dignity."

Belaboring the Obvious

Among the most persistent criticisms of social research is the "obviousness" of research findings. People ask, "Have social scientists discovered anything of significance that wasn't already known?" Researchers themselves question what can be said that is not trivial. Merton responds to the problem this way:

Should his [the sociologist's] inquiry only confirm what has been widely assumed . . . he will of course be charged with "laboring the obvious". . . If he ventures to examine socially implausible ideas that turn out to be untrue, he is a fool, wasting effort on a line of inquiry not worth pursuing in the first place. And finally if he should turn up implausible truths, he must be prepared to find himself regarded a charlatan. . . . Instances of each of these alternatives have occurred in the history of many sciences, but they would seem especially apt to occur in a discipline . . . that deals with matters about

which men have firm opinions presumably grounded in their own experience. (Merton, 1959: xv-xvii)

Although now many years old, Lazarsfeld's (1949) response to the charge of obviousness still remains one of the classics. Lazarsfeld said that to evaluate the proposition that social science is the "remorseless pursuit of proof of what everyone knew all along," one should consider the following obvious conclusions about the adjustment of servicemen to military life in World War II.

1. Better educated men showed more psychoneurotic symptoms than those with less education. (The mean instability of the intellectual as compared to the more impassive psychology of the man-on-the-street has often been commented on.)
2. Men from rural backgrounds were usually in better spirits during their army life than soldiers of city backgrounds. (After all, they were more accustomed to such hardships.)
3. Southern soldiers were better able to stand the climate in hot South Sea Islands than Northern soldiers. (Southerners are more accustomed to hot weather.)
4. White privates were more eager to become "noncoms" than Negroes. (Lack of ambition among Negroes is almost proverbial.)
5. Southern Negroes preferred Southern to Northern white officers. (The Southerners had a more fatherly attitude toward them.)
6. As long as the fighting continued, men were more eager to be returned to the States than they were after the German surrender. (You cannot blame people for not wanting to be killed.)

Why not just take such statements for granted rather than devoting expensive research effort to establish them? The interesting point is that research undertaken during the war refuted each of these statements. The fallacy of arguments about "common sense" and "what everybody knows" is that had the results of the investigation been mentioned first, they would probably have seemed as "obvious" as the erroneous statements given above.

Henshel (1971) also claims that the "obviousness" charge against the social sciences is historically naive. That is, many of the ideas derived from social research and theorizing have now become so commonplace that people forget that when they were introduced they were ignored or viewed with contempt—"the conflict that raged over their acceptance, their non-obvious character, are difficult to remember." His examples include the controversy in early criminology when social scientists presented data conflicting with early (common sense?) theories that attributed criminality to heredity or evolutionary "throwbacks." Henshel also notes that much of the early industrial research attributed worker productivity solely to external conditions such as lighting, or to worker states such as fatigue or monotony. The conclusions drawn from the Hawthorne studies that var-

ILLUSTRATION 15.2 Social science research is sometimes criticized because of the "obviousness" of its findings.



© 1973 by United Feature Syndicate, Inc: All rights reserved.

iations in worker productivity were largely a function of human factors as opposed to changes in the physical work environment are now so commonplace that we tend to forget that they challenged contemporary management theories and led to the establishment of new philosophies of industrial relations (for recent debates on this issue, see Franke and Kaul, 1978, and Bloombaum, 1983). In the area of intergroup relations, several well-founded racial and geographic theories, including the very notion that "race" as a genetic characteristic exists at all, were greatly weakened by the presentation of new sociological and anthropological data. Moreover, apparently it was not obvious to traditional political theorists that American voters did not fit the models of democratic citizenry held by classical democratic theorists until social scientists documented the fact. The fact that studies like *Voting* (Berelson, Lazarsfeld, and McPhee, 1954) have highlighted the discrepancies between voter characteristics and democratic theory makes possible the development of better theories of how democratic systems function.

The **principle of empiricism**—that observation (direct or indirect) is the ultimate arbiter of issues in the search for knowledge (Lundberg et al., 1968)—appears to be more soundly based than the strict reliance on common sense. If the main goal of the social scientist is the explanation and understanding of human behavior, then it follows that these goals are best achieved by use of the observational and analytical skills of empirical science.

Sociology and the other social sciences have made and will continue to make critical contributions in extending understanding of human behavior. Some findings of social research may conflict with current folklore,

but if such folklore cannot withstand the challenge of empirical investigation, it does not deserve to stand. The usage of social science information may be less effective than we might prefer, and we might not always agree with just how it is used; however, decision making that is informed by social science knowledge at least has the potential of achieving more positive outcomes than has decision making that is not informed by such knowledge. Weiss, describing the use of social science research in policy making, said:

It is not the kind of use most people have in mind when they hear the word. Not here the imminent decision, the single datum, the weighing of alternative options, and shazam! Officials apparently use social science as a general guide to reinforce their sense of the world and make sense of that part of it that is still unmapped or confusing. A bit of legitimation here, some ammunition for the political wars there, but a hearty dose of conceptual use to clarify the complexities of life. (Weiss, 1977: 17)

Greater understanding of human behavior continues to seem relevant to people's ability to live together harmoniously. If we are going to succeed in saving ourselves from ourselves, we need to better understand ourselves (Elms, 1972). Sociology and related disciplines continue to offer the potential for such understanding.

V. SOCIAL RESEARCH AND THE INDIVIDUAL CONSUMER

We conclude this chapter with what may seem to the reader a contradiction. In previous sections we discussed the use of social science research in policy making and in other areas that affect people's everyday lives. Our position has included an occasional lamentation that the contributions of the social researcher are not more widely recognized and used. We now consider the interpretation and use of research by individual consumers, and the perspective here is one of cautious skepticism.

The contradiction is more apparent than real. Social science knowledge is frequently underutilized in areas where it could have real programmatic impact and could contribute to an improved quality of life. However, there are other arenas where the findings of social research are abused, and it is frequently in such areas that the individual consumer of research is most directly affected.

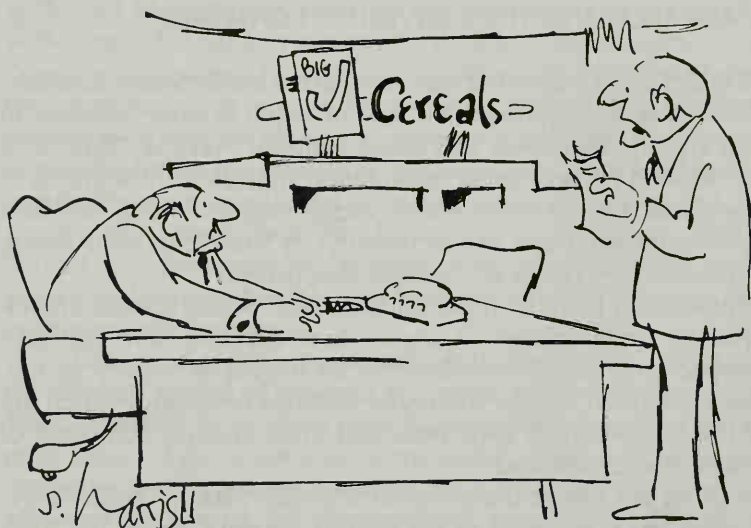
We live in an era where the status of scientific "truth" holds sway. "The talk of scientists is accepted as the nearest we can come to the truth about the world" (Weigert, 1981: 123). In many ways, science is both the magic and the religion of our time, and it has virtually replaced both of these realms in which "truth" was formerly grounded. Weigert notes that in a society organized around "magic," people had few alternatives but to

interpret life and resolve its problems in terms of magical truths. In a religious society, life was interpreted and problems solved by appeal to religious truths. However, in a scientific society, science is the final arbiter for truth. And that creates the problem that we pose in this section; like magic and religion, science can be abused or subverted to serve ends other than those of acquiring knowledge or improving the quality of human life.

We said earlier that most of us use scientific research findings regularly, either knowingly or unknowingly. However, the use to which we put such findings is often limited by our own training and imagination. If social research does anything for the individual consumer, it should provide a different perspective in evaluating and responding to the social world around us. Lord Alfred Whitehead (1925) once observed that "Familiar things happen, and mankind does not bother about them. It requires a very unusual mind to undertake the analysis of the obvious." The social scientist may be accused of discovering the obvious, but systematic analysis of that which seems obvious has greatly expanded our understanding of life.

At the same time, expecting too much is likely to generate disappointment rather than satisfaction. Those who look to the social sciences for immediate solutions to pressing social problems—crime, poverty, war,

ILLUSTRATION 15.3 Social science research is used to improve the quality of life of individual consumers.



"The report isn't all bad. They say that, eaten with warm milk, the box could provide some nutrition."

ethnic prejudice, the breakdown of the family, etc.—will find that social research offers relatively few tried and tested solutions. “Whatever importance the field may have lies less in its immediate than in its ultimate relevance, in its potential for yielding knowledge that will someday help provide lasting solutions to our chronically acute social problems” (Elms, 1972: 3).

In previous chapters we identified many parallels between the methods of the social sciences and the natural sciences, but there remain some important differences:

Natural scientists study objects and animals which do not have knowledge about their own world which itself makes up the core of that world, so such scientists apply any concepts or constructs they find effective. The electrons really do not care. Sociologists and others who study human life, however, confront a reality which is already someone else’s knowledge about the very subject matter under investigation, namely, human life. Humans shape their own reality and talk back to those who investigate it. (Weigert, 1981: 21)

Or, as another observer has noted: “Assembling bits and pieces about rats and ‘basic human behaviors’ doesn’t produce anything resembling complex social behavior, and it never will” (Elms, 1972: 7).

Some Additional Cautions

There are several other things that a potential user of research should understand. First, nearly everything can be admissible as evidence in research or in interpretation of research, but some evidence is more credible than others (Deutscher, 1973). C. Wright Mills (1940) argued for the use of a process he called **discounting**. That is, before interpreting data, one should consider such questions as the conditions under which the study was conducted, who did the study, who sponsored or provided the funding for the research, and so on. When a project is understood in its proper context, one has a better basis for evaluating the credibility that its findings and conclusions deserve. One bit of evidence might be used with a good deal of confidence while another might be discounted. However, it is important that we discount and interpret rather than merely discard:

We do not discard reports merely because of biases or flaws of one sort or another. . . . It is all presented by men who have some sort of stake in the matters of which they write, who are located somewhere in their own society (and tend to see the world from that perspective), and whose work is more or less open to methodological criticism. (Deutscher, 1973: 5)

Deutscher further proposes that in considering research evidence, we always employ what he calls a **double screen**. The two elements of this double screen are methodological and conceptual. That is, the reader first

evaluates a piece of work in terms of how well it was done, how reliable its methods were, and how technically competent it was; if it passes muster in this regard, one then evaluates its meaning and implications or its "conceptual adequacy." Only those things that are sound, methodologically and conceptually, deserve to be seen as credible evidence.

The methodological screen In the preceding chapters we stressed the importance of selecting tools appropriate to the research task. There is a broad range of available tools, and the research consumer should ask whether the right tool has been used. Although a problem can be approached in many different ways, some approaches are better suited than others to a given research context. It is critical that the researcher be aware of the options and their relative advantages; the research consumer is better prepared to interpret or understand if he or she also has some notion of the problems and advantages of various methods. Deutscher (1973) comments that social scientists sometimes tend to try to chop down trees with shovels because we have no axes but our shovels are good. Researchers may "force" a particular method, even if it doesn't fit, because they know that method well.

The critical question is, did the researcher use the method that was most appropriate for the problem at hand? A single, in-depth interview provides little basis for generalizing to a large population, and survey research is not the most effective technique for testing causal hypotheses. Returning to Deutscher's (1973: 41) description of the issue, ". . . we may have been learning a great deal about how to pursue an incorrect course with a maximum of precision."

A related question is whether the method used was used correctly. Each of the previous chapters provides some guidelines for deciding whether an appropriate method was properly implemented.

The conceptual screen Assessing the conceptual adequacy of a research report is often complex. Concepts are a key part of the entire research process. "Since science strives to achieve accuracy, every field of scientific endeavor develops a continuously refined set of concepts which, to the initiated, mean the same thing at all times under stated conditions. Thus it is imperative at the outset of any research effort to define clearly concepts . . . that will be employed" (Lastrucci, 1963: 77). Yet, conceptual clarity is often a critical problem in research and has a significant role in determining the credibility one should attribute to a set of findings or interpretations. "Vague ideas will lead to inadequate and uninterpretable research findings" (Labovitz and Hagedorn, 1971: 16).

Conceptual adequacy depends on the skills and purposes of the individual scientist. In fact, no method can ever be more accurate than the person who employs it. "So, since scientific method is an instrument of human reasoning, the reasoning processes themselves should be checked

for possible sources of error just as are all other tools or instruments" (Lastrucci, 1963: 86-87).

Several years ago Lastrucci developed an excellent summary of the most common errors in the employment of concepts. Among them are the following:

1. Incompleteness. The author's argument may employ either selected instances (to the exclusion of other relevant instances), or card stacking (the technique of amassing only supportive arguments while ignoring contradictory ones).
2. Preference for the familiar. Lastrucci notes that "whether the error takes the form of belief through simple and sheer repetition, or whether it takes the form of a slogan or other cliché, most persons gullibly accept as true that which is merely familiar."
3. The use of irrelevancies. This problem involves the use of some form of argument or proof that is logically unrelated to the propositions that are actually being tested in the research.
4. Incorrect deductions. This can take the form of overgeneralization, oversimplifying, or forcing one's findings into false categories.

For the consumer of research, such conceptual problems may be evident in advertising. To some degree, all four are apparent in the discussion of advertising for common pain relievers given in Illustration 15-4.

ILLUSTRATION 15.4

Try the following test: Which brand of headache remedy do you use? Why? Now check your answer with the following facts.

Brand A: Perhaps you use the most widely promoted brand of plain aspirin because you have heard that it is "100% pure aspirin" and that "government tests have proved that no pain reliever is stronger or more effective than Brand A." The makers of Brand A are quite right: their brand is all aspirin, and government tests did show that no pain reliever was stronger or more effective. It is also true, however, that no pain reliever in the tests was shown to be weaker or less effective than Brand A. That is, all brands were equally strong and effective.

In fact, of course, there is a difference among headache remedies: price! Currently Brand A is selling at my neighborhood drugstore for 98¢ for a 100-tablet bottle, whereas several other national and local brands are selling for 19¢ per 100 tablets. Medical consultants to Consumers Union, the world's largest non-profit consumer testing organization, continue to emphasize year after year that "you will get as good relief from common pains and fever as is available without a prescription if you buy the least expensive U.S.P. aspirin your store sells. . . . The only significant difference among brands of aspirin plain or buffered is price."

Brand B: Perhaps you use buffered aspirin because you have heard that it will not upset your stomach as aspirin might or because you have been persuaded by the commercials for Brand B (the best-known brand of buffered aspirin) that this product works "twice as fast as aspirin." But the government tests showed that there was no significant difference between plain and

ILLUSTRATION 15.4 *continued*

buffered aspirin (Brand B) in speed or in the incidence of stomach upset. But if you still want buffered aspirin, you can get it for about 25¢ per 100 tablets by buying a brand whose maker is not spending a fortune on national television advertising. Brand B at my neighborhood drugstore commands a price of \$1.49 per 100 tablets.

Brands C and D: At the time of the government tests, some headache remedies, including two of those tested, combined aspirin with phenacetin and caffeine. Accordingly, these are called APC tablets. Phenacetin has about the same effectiveness as aspirin in relieving pain and depressing fever, but it is not as good for suppressing the inflammation caused by such ailments as arthritis. As for caffeine, it is not a pain reliever, and there is no reliable evidence that it enhances the effect of aspirin either. Thus, it is not surprising that the government study found APC tablets to be no more effective than plain aspirin. The study did find, however, that APC tablets upset the stomach "with significantly greater frequency than any of the other products tested."

Since that time, of the brands tested, Brand C has replaced phenacetin with other compounds, retained the aspirin and caffeine, and raised its price—it is now \$1.59 per 100 tablets at my drugstore. The other combination tablet tested in the government study, Brand D, has simply had the phenacetin removed from it, leaving it with aspirin and caffeine. But this solution to the medical problem raised a public relations problem for the makers of Brand D. They had reportedly spent \$86,400,000 on a television commercial which showed three dishes of ingredients while a narrator said that Brand D contained "not one, not two, but a combination of medically proven ingredients." When the number of ingredients dwindled to an unimpressive two, it seemed a shame to abandon a sales pitch effective enough to have earned more money than the movie "Gone with the Wind." So at first the commercial kept the three dishes of ingredients, but moved them from the foreground of the television screen back into the package. After a few more months, the three dishes were deleted from the commercial but "combination of ingredients" was still heard. The next version had three dotted lines and a corresponding narration that Brand D contains (1) the pain reliever doctors recommend most (that's aspirin, of course), (2) plus more of that pain reliever (more aspirin), (3) plus the strength of another ingredient (caffeine). Thus 2 was made to equal 3.

The test is over. How did you do?

Daryl J. Bem, *Beliefs, Attitudes, and Human Affairs*. © 1970 by Wadsworth Pub. Co., Inc. Reprinted by permission of Brooks/Cole Pub. Co., Monterey, Calif.

The Business of Science

Finally, the consumer of research must distinguish between the content of science and the method of science (Lastrucci, 1963). The *content* of science constantly changes. For example, early researchers who studied

the concept "attitude" in social psychology assumed something approximating a one-to-one relationship between a person's attitudes and that person's behavior. This assumption has gradually yielded to the more complex view that the attitude-behavior linkage is a contingent one—that is, the strength of the link between an attitude and a behavior depends on many other characteristics of the actor and of the situation in which the behavior occurs. Therefore, the content of "science" on this topic has changed from an initial assumption that attitude was directly manifest in behavior to something like "attitude is related to behavior in the following ways under the following sets of conditions." The content of knowledge in any category of science changes as we accumulate more information.

On the other hand, the *method* of science is fairly constant: in its variety of forms and applications, such values as objectivity, logic, systematic analysis, and the accumulation of valid and reliable information continue to be emphasized. The various "methods" discussed in previous chapters are essentially "techniques" (Lastrucci, 1963) for doing a variety of research tasks. The selection of one over another depends upon the particular research problem. If a problem is clearly defined conceptually, appropriate methods are applied, and the analysis is systematic and competent, then the user or consumer of research findings may place considerable confidence in the findings.

VI. SUMMARY

Most chapters of this book have been designed to teach the reader how to do social science research. However, they have also been written to teach the reader how to *consume* research that has been done by others. Most readers will not go on to careers in research, but few will be able to avoid having contact with social science research in their daily lives. Such contact may occur via scientific public opinion polls, questionnaires that arrive in the mail, invitations to evaluate the effectiveness of a political candidate or a new school curriculum, and so on. Furthermore, everyone "consumes" research as he or she is influenced by the printed and electronic media, and many day-to-day decisions are affected by inputs from research of one kind or another.

In recent years social science research has had an ever-broadening array of applications. It has been used in law, medicine, social welfare, business management, foreign policy, urban planning, architecture, and many other disciplines. The role of the social science researcher in these varied fields of application has usually involved one or more of three activities: (1) sharing with a client the knowledge a researcher has gleaned from reviewing and summarizing social science research, (2) doing a special study, or (3) sensitizing clients or selected publics to perspectives on rel-

evant topics, such as ethnic or gender relationships, that are supported by research but not accepted as "common sense" definitions by people generally.

Social science research has also become increasingly important in public policy making. Its expanded role is reflected in social science involvement in major national commissions such as the National Advisory Commission on Civil Disorders and the Violence Commission. Research by social scientists has also received broad policy consideration, as in the instance of the well-known Coleman Report. Of special importance in social science research as a part of public policy making is the National Environmental Policy Act, which required social science research as part of environmental impact assessment.

Several barriers, however, still inhibit greater use of the social sciences in policy making. Among the most important of these are: (1) problems inherent in the research enterprise itself, including the tendency of social scientists to remain detached from society, to identify abstract impersonal causes for outcomes, to deal in generalities rather than specific problems, or to produce concepts difficult to apply in real-life situations; (2) the controversial nature of some social science research findings that threaten the values or practices of some interest groups; (3) problems with prediction, although we noted that in highly controlled or structured systems, the predictive ability of the social sciences is comparable to that in the physical sciences; and (4) the stereotype that social scientists belabor the obvious, despite extensive evidence to the contrary.

Broader application of social research in the life of the individual can also be noted. Modern society places great emphasis on science and its accomplishments. Indeed, individuals must exercise critical reasoning and caution lest they be misled or deceived by their reliance on scientific "evidence." The application of a "double screen" in interpreting research findings can help one to avoid some mistakes. The first part of the screen is methodological: did the research use methods that were most appropriate to the problem at hand? The second part of the screen is conceptual: were the variables defined appropriately and linked to conceptual frameworks such that the research findings are readily interpretable? Consumers of research—and that includes every citizen in modern societies—must exercise cautious skepticism in their acceptance and interpretation of research findings.

REFERENCES

- BEHAVIOR AND SOCIAL SCIENCES SURVEY COMMITTEE, 1970. *The Behavioral and Social Sciences: Outlook and Needs*. Washington, D.C.: National Academy of Science.

- BEM, DARYL J., 1970. *Beliefs, Attitudes, and Human Affairs*. Belmont, Cal.: Brooks/Cole.
- BERELSON, BERNARD R., PAUL F. LAZARSFELD, and WILLIAM N. MCPHEE, 1954. *Voting: A Study of Opinion Formulation in a Presidential Campaign*. Chicago: University of Chicago Press.
- BERGER, PETER L., 1963. *Invitation to Sociology: A Humanistic Perspective*. New York: Doubleday.
- BLOOMBAUM, MILTON, 1983. "The Hawthorne Experiments: A Critique and Reanalysis of the First Statistical Interpretation by Franke and Kaul." *Sociological Perspectives* 26 (January): 71-88.
- CARLSMITH, J. MERRILL, PHOEBE C. ELLSWORTH, and ELLIOT ARONSON, 1976. *Methods of Research in Social Psychology*. Reading, Mass.: Addison-Wesley.
- COLEMAN, JAMES S., 1975. "Racial Segregation in the Schools: New Research with New Policy Implications." *Phi Delta Kappan* (October): 75-78.
- COLEMAN, JAMES S., et al., 1966. *Equality of Educational Opportunity*. Washington, D.C.: U.S. Government Printing Office.
- DEUTSCHER, IRWIN, 1973. *What We Say/What We Do: Sentiments and Acts*. Glenview, Ill.: Scott, Foresman.
- ELMS, ALAN C., 1972. *Social Psychology and Social Relevance*. Boston: Little, Brown.
- FRANKE, R. H., and J. D. KAUL, 1978. "The Hawthorne Experiments: First Statistical Interpretation." *American Sociological Review* 43 (October): 623-643.
- GANS, HERBERT J., 1971. "Social Science for Social Policy." Pp. 13-33 in Irving Louis Horowitz (ed.), *The Use and Abuse of Social Science*. New Brunswick, N.J.: Transaction Books.
- GREEN, ROBERT L., and THOMAS F. PETTIGREW, 1976. "Urban Desegregation and White Flight: A Response to Coleman." *Phi Delta Kappan* (February): 399-402.
- HENSHEL, RICHARD L., 1971. "Sociology and Prediction." *The American Sociologist* 6 (August): 213-220.
- HOROWITZ, IRVING LOUIS (ed.), 1971. *The Use and Abuse of Social Science*. New Brunswick, N.J.: Transaction Books.
- LABOVITZ, SANFORD, and ROBERT HAGEDORN, 1971. *Introduction to Social Research*. New York: McGraw-Hill.
- LASTRUCCI, CARLO L., 1963. *The Scientific Approach: Basic Principles of the Scientific Method*. Cambridge, Mass.: Schenkman.
- LAZARSFELD, PAUL F., 1949. "The American Soldier—An Expository Review." *Public Opinion Quarterly* 13 (Fall): 377-404.
- LAZARSFELD, PAUL F., WILLIAM H. SEWELL, and HAROLD L. WILENSKY (eds.), 1967. *The Uses of Sociology*. New York: Basic Books.
- LUNDBERG, GEORGE A., et al., 1968. *Sociology* (4th ed.). New York: Harper & Row.
- MERTON, ROBERT K., 1959. "Notes on Problem-finding in Sociology." Pp. ix-xxxiv in Robert K. Merton et al. (eds.), *Sociology Today*. New York: Basic Books.
- MILLS, C. WRIGHT, 1940. "Methodological Consequences of the Sociology of Knowledge." *American Journal of Sociology* 46 (November): 316-330.
- NATIONAL ADVISORY COMMISSION ON CIVIL DISORDERS, 1968. *Report of the National Advisory Commission on Civil Disorders*. New York: Bantam Books.
- NATIONAL COMMISSION ON THE CAUSES AND PREVENTION OF VIOLENCE, 1969. *Violence in America: Report to the National Commission on the Causes and Prevention of Violence*. New York: Signet Books.
- SIBLEY, ELBRIDGE, 1971. "Scientific Sociology at Bay?" *The American Sociologist* 6 (June): 13-17.
- TILLY, CHARLES, 1969. "Collective Violence in European Perspective." Pp.

- 4-42 in *Violence in America, Report to the National Commission on the Causes and Prevention of Violence*. New York: Signet Books.
- WEICK, KARL E., 1965. "Laboratory Experimentation with Organizations." Pp. 194-206 in J. G. March (ed.), *Handbook of Organizations*. Chicago: Rand McNally.
- WEIGERT, ANDREW J., 1981. *Sociology of Everyday Life*. New York: Longman.
- WEISS, CAROL H., 1977. *Using Social Research in Public Policy Making*. Lexington, Mass.: Lexington Books.
- WEISS, CAROL H., and MICHAEL J. BUCUVALES, 1980. "Truth Tests and Utility Tests: Decision-Makers' Frames of Reference for Social Science Research." *American Sociological Review* 45 (April): 302-313.
- WHITEHEAD, ALFRED NORTH, 1925. *Science and the Modern World*. New York: Macmillan.
- ZELDITCH, MORRIS, JR. n.d. "Can You Really Study an Army in the Laboratory?" Unpublished manuscript. Stanford University.

glossary

analysis the process of studying the nature of events, cultures, practices or other social entities by identifying component parts and how they relate to each other and to the larger entities within which they occur.

analysis of variance a test of statistical significance that is used when the dependent variable is measured intervally and there are multiple nominal categories.

applied research research oriented to finding solutions to specific social problems.

area sample, area probability sample multi-staged cluster sample that starts with a sample of larger geographical areas as sampling units like counties and then shifts to smaller units like towns, census tracts, blocks, and eventually households.

attitudinal set tendency to respond in a given way or to have preconceived notions which bias perceptions, values, beliefs or interpretations of events in a consistent way.

baseline profile a comprehensive description of the study area prior to the introduction of a development of some type. Typically includes a discussion of how many people live in an area, their characteristics, and so on.

baseline projection designed to predict how an area is likely to change *without* a proposed development.

basic research research directed toward advancing scientific knowledge for its own sake.

between-method triangulation use of different research techniques to study the same problem.

bi-modal distribution in a frequency distribution, the score that occurs most frequently is the mode; in a bi-modal distribution, two scores rather than one appear most frequently.

category a statement that describes a given class of phenomena into which observational behavior may be coded. In observational or secondary research, the researcher develops a series of categories into which the raw data or observations are to be clustered.

causality the relation of cause and effect; as a methodological approach, the idea that one event is the net product or result of other specifiable, knowable events or conditions.

chi-square a test of statistical significance that is commonly used for ordinal level data.

closed-response item a question included in an interview or questionnaire that forces the respondent to choose one or more responses that the researcher provides.

cluster sample the population is divided into clusters of sampling units (households, families, classrooms, schools, etc.), a sampling of clusters is selected and all the units in the selected clusters are studied.

codebook systematic instructions for translating data from one form or language to another; ideally provides for each variable or item of information in the data-collection instrument an unambiguous guide for translating any response into an interpretable numerical system.

coding the reading of, listening to, or observing of data and placing the observation in specific categories.

coefficient of determination measures the percentage of variance in a dependent variable that is determined by the independent variable.

coefficient of reliability a statistical computation assessing the level of reliability between two or more coders.

coefficients of association measures of association that are applied when one's data are either nominal or ordinal.

comparative research using contrasts between different cultures, groups, systems or other social categories as a primary method of analysis and interpretation.

concept a general notion or idea; a mental device which presumably assists its user in dividing perceptions into meaningful categories; an abstraction used to analyze or interpret observations.

connotative definition a secondary or implicit definition of a term or concept; a "hidden" meaning a term may carry for some who use it.

content analysis the study of communications to describe social behavior or to test hypotheses about it.

contingency table tabular presentation of numerical data in which knowledge of totals and digits in cells for one variable determines cell frequencies for the values of a second variable, usually applied to 4 cell (2x2) tables.

continuum continuous series of fine gradations in quality or quantity of a theoretical or measurable variable.

convenience sample sampling units that are convenient or readily available to the researcher are selected to represent a population.

correlation analysis used to determine whether two or more variables are related; that is, whether change in a dependent variable is affected by change in an independent variable.

correlation coefficient measure of association between variables that is applied when one's data are interval or ratio.

cover letter letter explaining purpose of mail questionnaire study, providing instructions how to complete questionnaire and appeal for participation.

cross-sectional measurement at a single point in time which obtains representation of elements or characteristics present in a unit or population at that time, but does not reflect characteristics before or after the point or period of measurement.

cross-tabulation analytical technique which involves sorting of unit characteristics along two or more variables, such that each cell in the table represents a different combination of the categories of each variable, and the entire table includes all possible combinations for the particular category systems being used. For example, a cross-tabulation of 2 two-category variables produces a 4-cell table; of 2 three-category variables a 9-cell table, and so on.

cumulative formed by the addition of successive parts or elements; e.g., a cumulative distribution at any point includes the total of all frequencies or percentages at previous (lower) points or categories.

cumulative scaling an attitude scaling procedure developed by Louis Guttman. This procedure is commonly referred to as scalogram analysis and seeks to develop a set of items for attitude measurement that will be unidimensional.

data analysis the study of measurements, accounts, or other bits of information to identify parts, patterns, and processes; see "analysis."

data entry mechanical or electronic transmission of units of information into cumulative combinations of unit data in form more accessible to electronics or other mechanical analysis than were the original units.

demographic items questions included in an interview or questionnaire that obtain information about social background characteristics such as age, sex, race, marital status, education, and so on.

dependent variable the phenomenon the researcher is interested in describing, explaining, or predicting.

descriptive analysis analysis in which the primary objective is identifying and labelling characteristics and components of the subject or population under study rather than with processes and interrelationships.

descriptive statistics measures such as percentages, means, and standard deviations that are computed to describe the characteristics of one's sample.

difference of means a statistical test used to determine whether an observed difference between two independently drawn samples is real or could have occurred by chance.

difference of proportions a special case of the difference of means test. Scores are translated into proportions and one uses the test to determine if the difference between the proportions is statistically significant.

discounting a procedure for evaluating the quality of research. Typically one should consider such questions as the conditions under which the study was conducted, sponsorship for the study, and so on.

disinterestedness characteristic of the "ideal" scientist; a lack of personal stake in or commitment to any potential outcome of a study or experiment.

disproportionate stratified sample the population is divided in such a way that units have unequal probabilities of being included in the sample.

double-barreled question a question in an interview or questionnaire that asks two independent questions and the researcher cannot be sure which was answered.

double screen a procedure for evaluating a research study by first considering the methodological accuracy of the study and then by evaluating its conceptual adequacy and clarity.

effort evaluation in evaluation research, effort evaluation is undertaken to assess the quantity and quality of activity evident in program implementation.

elegance a characteristic of scientific explanation or research design; denotes simplicity of explanation, use of as few components, concepts, variables, etc. as possible to provide sufficient explanation or evidence.

empirical based on experience or evidence; empirical methods involve testing and measurement rather than hypothetical argument or theorizing.

environmental impact assessment a broad-scale assessment of the impacts of a development or proposed development. An environmental impact assessment typically includes an analysis of the impacts a development will have on the local ecosystem, including the economy, animal and plant life, air and water quality, agricultural production, as well as the quality of life of local residents.

experiment a research technique where one or more independent variables are manipulated by the researcher under controlled conditions to determine changes caused in the dependent variable.

experimental mortality subjects dropping out of an experiment before it is completed.

explained variance the amount of variance in one variable that is explained by another variable.

exploratory interview loosely structured, flexible interview whose objective is identifying the general context and possible avenues for more focused questioning;

may refer to an initial interview with a respondent who will later be subjected to more formal questioning, or to a series of interviews with different respondents aimed at learning enough to design an appropriate research project; may also be the only research technique used in inquiries where a high degree of quantitative precision is unnecessary.

external validity the degree to which the findings of a study can be generalized to other populations and other settings.

face-to-face interview data collection by personal encounter, in which interviewer is in the presence of the respondent in contrast, for example, to the telephone interview.

factor analysis a procedure that is commonly used to reduce large sets of indicators to one or a few composite scales. A technique for data reduction.

field experiment an experiment that is conducted in a real social setting such as a school, factory, mental hospital, prison, church, club or home.

field log a device for recording observations made during observational research. A field log might be a detailed diary kept by a participant observer, a set of notes that is prepared at the end of each observational period, or information recorded on a pre-prepared recording schedule.

follow-up contact repeated attempts to contact potential respondents and encourage them to participate in the research project.

formal theory a framework of concepts and variables purporting to describe some aspect of reality by the set of interrelated elements, the elements and their relationships being systematically and precisely defined.

frequency distribution a listing of numbers of units of a given population along each of the categories or points of the scale being used to measure a characteristic of the population.

gamma a measure of association used for ordinal level data.

Hawthorne effect changes in a dependent variable due to the knowledge and attendant modification of attitudes and behavior of subjects in an experiment that they are being studied; the "guinea pig" effect.

Heisenberg principle one cannot observe something without changing it.

history in experiments, a source of variance in the dependent variable due to uncontrolled events preceding measurement or intervening between measurements.

hypothesis statement of possible relationship between variables in a form suitable for empirical test.

hypothesis-testing a type of research design in which previously articulated hypotheses are assessed empirically, in contrast to exploratory or descriptive research which typically does not test formal hypotheses.

impact a change in a current or established economic, demographic, or social condition that is caused by the introduction of a development of some type.

impact analysis examines the differences that occur between the baseline projections and the impact projections in order to determine the kinds of changes that are likely to be associated with a proposed development.

impact assessment examines the outcome of program implementation.

impact mitigation the identification and application of a set of procedures or measures designed to minimize the potential negative impacts or maximize the potential positive impacts that are associated with a proposed development.

impact monitoring typically done after the completion of a social impact assess-

ment and involves the task of continuing to collect data while a project is being implemented in order to determine if projected changes do, in fact, occur.

impact projection predicts the effect that the project will have on a community beyond those changes that will already be associated with baseline projections.

independent variable phenomenon that is used to explain or predict the dependent variable.

indexing a combination of a series of single measures or indicators into a single measure of a variable.

inferential statistics measures that are computed to infer properties of a population from information that is obtained on a sample that is drawn randomly from that population.

informed consent a research subject decides to participate in a study fully understanding the nature of the study and the risks, and free from force, fraud, deceit, or duress.

instrument a general name for both an interview form, a questionnaire, or other data-collection device.

instrument decay inaccurate observations because observers have become biased, lazy, or fatigued, or because a mechanical device for data collection has worn out through misuse or overuse.

internal validity the degree to which extraneous variables are controlled so that the researcher is confident that change in the independent variable(s) caused the changes in the dependent variable.

interpretation assigning meaning to findings.

interval scale level of measurement in which units are equally spaced or have equal values such that arithmetic and algebraic manipulation of scale scores is appropriate.

interview a survey research technique where the researcher asks the respondent questions and records the answers.

interview guide the exact questions the interviewer wants to ask and the order they are presented in. Sometimes called an interview schedule.

investigator triangulation replication of research project or tests of hypotheses by different scientists who use the same methods.

laboratory experiment an experiment conducted in a laboratory allowing the experimenter control over relevant variables and conditions.

lambda a measure of association used for nominal level data.

longitudinal measurement across time; also refers to studies of individuals or other units in which measurements of the same unit are repeated at various times over the course of the study.

mail questionnaire survey a study where the questionnaires are delivered to the respondents and returned to the researcher via the mail.

marginals the sum of frequency distributions, or of units, represented in categories, as in a cross tabulation table; row and column totals are examples of marginals, in that they are the subtotals that appear in the margins of the tables.

maturation biological, psychological and emotional growth or change in respondents or subjects over time that may influence the dependent variable in an experiment.

mean the arithmetic average computed by summing all scores and dividing that sum by the number of individuals or units tested.

measurement the assigning of a value to observations so that comparisons can be made.

measures of association statistics that assess the strength of a relationship between variables or the amount of variation in the dependent variable that can be explained by the independent variable.

mechanical recording device a mechanical aid for recording observations in an observational study. Such devices might include audio recordings, video recordings, and motion pictures.

median that point or score which precisely divides a frequency distribution into two equal parts; that score which has the same number of scores above it as below it.

method of equal appearing intervals an attitude scaling procedure developed by Thurstone assumed to have interval level properties.

method of summated ratings an attitude scaling procedure developed by Rensis Likert having ordinal level properties.

methodological triangulation replication of research projects or tests of hypotheses using different research procedures than those used in an initial test.

mode the most frequent score in a frequency distribution.

multi-method approaches research designs which use a variety of research techniques (i.e., methodological triangulation) to test an hypothesis or study a topic.

multiple correlation a statistical technique used to measure the total variation in a dependent variable which can be explained by a series of independent variables acting together.

multiple indicators items presumably measuring the same variable, or at least aspects of the same variable, used together to provide greater assurance that the underlying variable was in fact well-represented in the results.

multiple operationism use of a variety of data collection and analytical techniques to allow for the weaknesses of the individual methods and giving greater confidence in similar findings secured in different ways.

multiple regression a statistical procedure that is used to predict the value of a dependent variable (X_1) from a series of independent variables (X_2 , X_3 , etc.).

multiple regression analysis application of multiple regression techniques to a body of data.

multivariate analysis application of a variety of methods which consider the relative importance of many independent and intervening variables on a dependent variable.

NEPA National Environmental Policy Act passed by Congress in 1969, which established the requirement that comprehensive, systematic evaluations of the effects of developments on the natural and human environments be conducted for all projects conducted by federal agencies or affecting lands or other entities over which the federal government has regulatory or management responsibility.

natural experiment an experiment where the researcher obtains baseline levels on the independent and dependent variables and then waits for the independent variable to occur or change naturally. Independent variables such as floods, earthquakes, and factory closing are generally studied in natural experiments.

nominal measurement the assignment of numbers or scores to categories of a variable for identification purposes, when no assumption of rank or quantity are intended; such scores may not be manipulated algebraically or arithmetically.

non-random sample sampling units are selected to represent a population in a way so that units have unequal probabilities of being included in the sample.

null hypothesis the hypothesis that is actually tested as opposed to the research hypothesis. Typically, the null hypothesis is stated in negative terms—that is, that there is no relationship between the variables.

OMB Circular A-40 a set of governmental regulations that specify when and how one can use questionnaires and interviews in conducting research that is sponsored by governmental agencies. Specifically, the regulation states that one cannot use a questionnaire that asks the same set of questions to more than nine persons without first obtaining OMB clearance.

obtrusive observation measurement or observation which is conducted in ways that make the persons or social units under study aware of the research process going on; the more obtrusive the observation, the more it “bothers” the subjects of observation.

official goal goal of a program that is stated in official reports or in public statements by program supporters and leaders.

open-ended item a question included in an interview or questionnaire that allows the respondent to answer in his or her own words.

operational definition a specific set of instructions explaining how a variable is measured.

operative goal goal that an organization actually attempts to achieve. Operative goals may or may not be consistent with official goals.

ordinal measurement level of measurement which scores units as “higher than” or “lower than” other units but does not specify the size of the interval of quality or quantity separating them; ordinary arithmetic and algebraic manipulation of ordinal data are inappropriate, but a set of appropriate mathematical procedures for manipulation of ordinal data have been created.

originating question the first component in the progressive formulation of a scientific problem, namely the statement of what one wants to know.

parameter characteristics of a population as contrasted with characteristics of a sample drawn from that population.

parsimony economy of concept, design, or procedure; the principle of “as much as necessary, but no more” in theory or procedure.

partial correlation a statistical procedure that is used to reflect the relationship between a dependent variable and each of several independent variables, while eliminating any tendency for the remaining independent variables to affect the relationship.

participant observation the researcher participating in the normal activities of the subjects he wishes to study in order to observe and feel the world from the subject’s perspective.

personal delivery technique a study where the questionnaires are delivered to the respondents by the researcher or an assistant.

policy problem issue in applied research related to the need of some organization or government to initiate, amend, or revoke formal procedures or requirements.

population the total group the researcher wishes to study. Sometimes called the universe.

precoded items in a questionnaire or interview schedule for which response categories have corresponding numbers printed on the instrument to facilitate coding and data entry.

pretest the testing of an interview guide, questionnaire or data collection technique on respondents not in the sample before the data collection with the selected sample is accomplished.

primary data collection the collection of data from subjects or respondents as compared to using data already collected by someone else (secondary data).

principle of empiricism the principle that observation, either direct or indirect, is the ultimate arbiter of issues in the search for knowledge.

probability a determination of the likelihood that something occurs by chance, as in assessing whether an observation made from a sample exists in the population from which the sample was selected.

probe to ask for greater detail about a response to a question in an interview or questionnaire.

process evaluation in evaluation research, process evaluation is used to examine the actual operation of the program and to see how the internal dynamics operate.

program planning a type of evaluation that focuses on identifying the program target population, determining how to reach that population, assessing whether the program is really needed, and so on.

program monitoring a type of evaluation that examines how a target population is being served, whether appropriate staff have been hired, and so on.

project description in an impact assessment, the project description describes the characteristics of a proposed development such as the number of new workers that will be required, the length of their stay in the community, and so on.

projective personality test social psychological measurement in which the subject reacts to a word, object, or form in a manner that presumably reveals or "projects" useful information about his or her personality; the Rorschach test is an example.

proportional reduction in error a property of certain measures of association that allows the researcher to determine the amount of reduction in error in predicting the dependent variable that can be accounted for by knowledge of the independent variable.

proposition a statement about the relationship between two or more concepts or variables in a scientific theory; may or may not be testable.

pure research research guided by a disinterested wish to know or commitment to the advancement of scientific knowledge, and which is conducted without the justification that the results will be useful in any practical or economic sense.

purposive sample researcher uses professional judgment or expertise to select sampling units that represent a population.

qualitative research research strategies that emphasize getting close to data, participation and experience as opposed to numerical counting of social behavior.

quantitative research research strategies that emphasize careful measurement of the social behavior being studied.

questionnaire a set of written questions that the respondent answers by himself or herself.

quota sample information about the population is used to determine how many (quotas) sampling units with specific characteristics should be selected to represent the population.

random sample the selection of sampling units from a population so that every unit has exactly the same probability of being included in the sample.

range the span of measurement intervals between the highest and lowest score in a frequency distribution.

ratio scale an interval scale anchored to an absolute zero point.

rationale the statement of reasons why a question is worth asking or a project worth doing.

rappro interpersonal harmony and empathy or understanding between researcher and subject or interviewer and respondent.

reactive effect that portion of the effects of manipulating an independent variable or doing a study which results from the experimental context as it was interpreted and reacted to by the subject; effects stemming from the intrusion of data collection.

regression analysis used to test for the impact of an independent variable on the dependent variable while statistically controlling for the effects of other independent variables.

reliability the consistency or stability of the measurement of a variable or the results of a study.

reputational approach a procedure used in key informant interviewing whereby potential interview subjects are identified by asking others to identify those who really know what is going on or who would be most highly informed about a particular topic or issue.

research design the systematic description of all stages of the research process; the sequence of activities necessary to answer the research question.

research proposal a formal statement to a funding agency or administrative authority containing a research question, justifying its significance, providing a rationale and research design for proceeding to study the question; usually involves statements of investigator capabilities and budget.

resource efficiency in evaluation research, relates to the efficiency with which resources available for the program are used. The typical question is whether an equally effective but less expensive program might be used to achieve program objectives.

response rate the percent of respondents in the sample who participated in the study.

restricted sample a sampling unit selected for a sample is not returned to the sampling frame before the next unit is selected.

retrospective questions about the past, usually in questionnaire or interview studies, which the respondent answers from present knowledge or memory; reconstruction of past events or experiences.

ritualism activities performed for their own sake or without regard to an objective which formerly was linked to the activities.

sample a set of sampling units selected for study; the results are generalized to the population they represent.

sampling frame a list of sampling units comprising a population from which a sample is selected.

sampling unit the individual person or element (classroom, work group, household, etc.) when combined with all other persons or elements make up the population to be studied.

scaling the combination of several measures into a single, composite variable as in attitude scaling.

schedule a device for recording observations or for asking questions and recording answers in an interview study.

science the process of studying and interpreting existence and observation and organization of principles, propositions, on a fact for which there is empirical support; also, the knowledge acquired by systematic, empirical inquiry.

scientific method method of assessing the validity of ideas about reality and existence by a systematic study and observation, combined with the recording of observations and how they were obtained so that resulting "facts" may be checked and modified by others.

screening questions a question included in an interview or questionnaire that identifies respondents with specific characteristics so they can be asked specialized questions.

secondary analysis the use of research materials by persons other than those who gathered them and/or for purposes different than the original project objectives.

selective perception the tendency in observational research to pay attention to certain events or occurrences at the expense of others that might be equally or more important and theoretically relevant to the study.

significance level tells us how confident we can be that the findings observed from a sample are applicable to the population from which the sample was drawn.

simple linear regression used to determine the degree to which one variable changes with a given change in another variable.

skewed distribution a distribution of scores or responses in which there are considerably more extreme cases in one direction than in the other.

snowball sample sampling units with relevant characteristics are selected and they in turn identify additional units with the same characteristics to be included in the sample.

snowballing a technique used in interviewing to identify a set of respondents. Each person interviewed is asked to identify additional respondents who could or should be contacted as part of the study. In this way, the pool of potential respondents continues to be expanded until the researcher has obtained the necessary number of interviews.

social impact assessment an applied form of social research that focuses on assessing the impacts of a development on a community or larger geographic region. Impacts that are typically assessed are demographic, community service, and quality of life changes that are wrought by the development.

software the programs, data, and routines for use in a computer, as distinguished from the physical components.

standard deviation a statistical measure of dispersion or variation of a series of scores around the mean.

standard normal distribution a symmetrical distribution in which cases or observations are distributed equally on either side of the mean.

statistic a characteristic of a sample that is used to estimate a parameter of a population.

statistical regression the tendency for extreme behavior in subjects to be replaced, as a general rule, by less dramatic, more normal behavior.

statistical significance refers to the probability of an observation or finding occurring by chance as opposed to being real.

stratified sample the population is divided into subgroups called strata and independent samples of each stratum are selected.

structured observation in observational research, a structured observation typically focuses on clearly defined aspects of behavior or subjects. These are identified and operationally defined before the researcher begins observing.

survey research a research technique that asks questions of a sample of respondents with a questionnaire or an interview.

systematic observation observation that is guided by predetermined decisions about what is to be observed, when and how observation is to occur, and so on.

systematic sample the sampling units on the sampling frame are numbered and every n th unit is selected for the sample. The first unit is randomly selected between 1 and N in the sampling frame.

telephone interview data collection using an interview guide over the telephone.

testing effect the possibility that an initial observation in testing may influence subsequent observations or test results independent of any independent variables.

tests of significance statistical tests that allow the researcher to determine whether or not a relationship is real or could have occurred simply by chance.

theory a set of interrelated hypotheses or propositions concerning some phenomenon. Most social science research is ultimately designed to test theory and the best research questions are often those that are derived from theory.

treatment evaluation in evaluation research, treatment evaluation is intended to document exactly what treatments were given and their effectiveness.

triangulation use of more than one method, or more than one project, to study a single problem.

unidimensionality a property commonly associated with scalogram analysis or cumulative scaling that assumes that an individual with a higher rank or score than another on the same set of attitude statements must also rank as high or higher on every statement in the set as the other individual.

unit of analysis the element of communication such as a word, paragraph, theme, or event that is coded into categories when doing content analysis.

universalism application of a single standard or set of standards to problems, issues, or individuals regardless of the status, rank, or power of the persons or units being studied, and without regard to their relationship to the investigator.

universe the total group the researcher wishes to study. Sometimes called the population.

unobtrusive measures or procedures which do not impinge upon persons or units being studied; if a method is truly unobtrusive the investigator does not influence, affect, interact with, or "bother" respondents or other units under study in any way.

unrestricted sample a sampling unit selected for a sample is returned to the sampling frame before the next unit is selected.

unstructured observation observation of naturally occurring events, usually as they happen. Typically the researcher has no control over the events and simply records them as they occur.

validity the degree to which an operational definition actually measures the variable it represents.

variables phenomena that are being studied. Usually they occur at different levels of intensity and thus the label variable.

verification subsequent tests of a finding sufficient to raise the probability of its accuracy to the point that other scientists can accept it as accurate.

within-method triangulation the repeated application of a single method to a research problem to assure that findings in an initial procedure were not biased in some way.

INDEX

- Aaronson, S., 398, 401
 ABA design (Time Series), 185-186
 behavior modification experiment, 194-197
 Absolute zero, 364
 Abstraction, of reality, 12-13
 Advertisements, content analysis source, 246
 Albrecht, D. E., 320, 322, 323, 332, 337
 Albrecht, S. L., 56, 136, 140, 153
 Alcorn, D. S., 136
 Alfred, R. H., 212-213
Alphabetical Index of Industries and Occupations (Census Bureau), 355-356
 American Institute of Public Opinion (Gallup Poll), 128
 Ames, N. L., 292
 Analysis, 14, 433
 observational data, 87-88
 Analysis and report:
 behavior modification experiment, 196-197
 content analysis, 250
 experimental research, 187
 guaranteed income experiment, 198, 201-202
 in homosexual behavior experiment, 234-235
 in Indian activism study, 254
 in prophetic failure experiment, 230
 qualitative research, 224
 in selection for obedience to authority experiment, 190
 in violence profile study, 252
 See also Writing about research.
 Analysis of variance, 372, 433
 Andreski, S., 395
 Applied research, 9, 27, 282-283, 433
 See also Evaluation research.
 Archives:
 catalogues of, 269
 secondary data, 266, 268-269
 Area probability sampling, 62-65, 433
 Arnold, C. R., 192
 Aronow, E., 243
 Aronson, E., 420
 Attitude scales, 384-386
 Attitudinal set, 109, 433
 Attkisson, C. C., 243
 Audience identification, in research writing, 399
 Auger, C. S., 319

 Background questions, questionnaires, 144
 Backstrom, C. H., 121, 123
 Bahr, H. M., 31, 56, 108, 110, 113, 160, 216, 241, 247, 273, 275-276
 Bahr, R. S., 246
 Bailey, K. D., 68, 260
 Bale, R. N., 107
 Bales, R. F., 80, 84
 Ball, J. C., 107, 108
 Barass, R., 391
 Barnett, J., 138
 Barrass, R., 397
 Barriers to research utilization:
 characteristics of social research and, 416-418
 controversial nature of, 418-419
 obviousness of findings, 420-423
 prediction problems, 419-420
 Barzun, J., 402, 403
 Baseline profile, 324, 328, 433
 Baseline projections, 324-325, 433
 Basic research, 282-283, 433

 Bechtol, B. E., 105
 Becker, H. S., 87, 104, 116, 211, 335
 Becker, W. C., 192
 Behavior factors:
 observation of, 81-82
 sampling for observation, 85-86
 Behavior modification experiment, 191-197
 Behavioral psychology, ABA design, 185-186
 Bellah, R. N., 218
 Bennett, C. A., 299
 Benney, M., 103
 Bensman, J., 105, 109, 110
 Berelson, B. R., 422
 Berger, P. L., 395, 419
 Beta coefficients, 383
 Between-method triangulation, 41, 433
 Bias:
 experimenter, 180
 questionnaires and, 138
 sources in interviewing, 112-115
 Bimodal distribution, 363, 433-434
Black Like Me (Griffin), 219-221, 223-224
 Blalock, H. M., 369, 378, 379, 381, 383
 Bloombaum, M., 422
 Bohnstedt, M., 304, 305, 307
 Boies, S. J., 403
 Bolitho, W., 66
 Booth, V., 396
 Boruch, R. F., 261, 268, 302
 Boulding, K. E., 4, 11, 13, 21, 316, 395
 Boutry, G. A., 393
 Bowker, L. H., 246
 Bradburn, N. M., 137, 144
 Branch, K., 318, 322, 333, 340
 Brandt, A. M., 16, 17, 18
 Breiver, B., 248
 Brekke, M., 109, 125, 138
 Bronowski, J., 22
 Bryant, F. B., 265
 Buckholdt, D., 93-95
 Bucuvalas, M. J., 411-412, 417
 Bugental, D., 91
 Bushnell, D., 93-95
 Buzzard, R. B., 401

 Campbell, B. A., 106
 Campbell, D. T., 170, 177, 178, 179, 180-181, 302
 Cannell, C. F., 103, 109, 110, 111-112, 113, 114-115, 118, 119
 Caplovitz, D., 373
 Caplow, T., 31, 35, 216, 241, 247, 395
 Cardon, P. F., 18
 Carlisle, E. F., 396
 Carlsmith, J. M., 420
 Cartoons, content analysis source, 246
 Cary, M. S., 81
 Category systems:
 use in content analysis, 248-249, 250
 for recording observations, 83-85
 Causability, 47, 434
 Cause and effect, multimethod approach, 244-245
 Census data, 328, 330
 Chadwick, B. A., 31, 56, 136, 140, 149, 153, 160, 179, 191, 216, 241, 242, 247, 275-276
 Chai, T. R., 244
 Chen, H. T., 296, 299
 Chi-square, 370-371, 434

- Christiansen, J. W., 246
- City directories, as sampling frames, 59
- Cleaning data, 346, 356-357
- Clinton, L., 275-276
- Closed-response items, 118-119, 434
- Cluster sample, 61-62, 434
- area probability sampling, 62-65
- Codebook, 346-353, 434
- Coder training and reliability checking:
- content analysis, 249-250
- in Indian activism study, 254
- in violence profile study, 252
- Coding, 346-356, 434
- code-building, 352-353
- Coefficient of determination, 379, 434
- Coefficient of reliability, 249-250, 434
- Coefficients of association, 374, 434
- Cognitive dissonance theory, 230
- Coleman, J. S., 44, 62, 147, 155, 157, 160, 168, 414-415
- College students, as experimental subjects, 171, 173
- Communication:
- components of, 240
- distortion of observation in, 13-14
- in interview studies, 115-116
- measurement and, 42-43
- See also* Content analysis.
- Communication source:
- content analysis, 248
- in Indian activism study, 253
- in violence profile study, 251
- Communism, 8
- Community impact analysis, sources of data, 329
- Community research, *See* Social Impact Assessment.
- Community service impacts, 320-321
- Comparative research, 33, 434
- Complex questions, 117
- Computers:
- content analysis, 247
- data analysis, 355-356
- data entry, 357
- error detection, 357
- questionnaire data analysis, 153
- Comstock, G. A., 241
- Concept, 14, 434
- Concepts:
- adequacy in research project, 426-427
- descriptive and operative, 14
- errors in use of, 427
- Condie, S. J., 29, 246
- Consumer of research, 423-429
- cautions to, 425-428
- content and method of science distinction, 428-429
- Content, of observation, 81-82
- Content analysis, 239-256, 434
- coder training, 249
- computer assisted, 247
- conducting content analysis, 248-250
- Indian activism study, 252-255
- sources of material, 245-247
- strengths and weaknesses, 243-245
- violence profile study, 250-252
- Content analysis procedures, 248-250
- analysis and report, 250
- communication source, 248
- operational definitions, 248-249
- statement of problem, 248
- training coders and checking reliability, 249-250
- Contingency tables, 368, 434
- Continuum, 43, 434
- Control groups:
- in pretest-posttest control group design, 184
- in static-group comparison design, 182-183
- in Solomon-Four Group design, 184-185
- Controversy, of social science research, 418-419
- Convenience sample, 65, 434
- Converse, J. M., 106, 109, 110, 111-112, 113
- Cook, S. W., 74, 79, 80, 86
- Cook, T. D., 287, 289, 290, 301
- Cooperud, R.H., 402
- Cordray, D. S., 261, 268
- Correlation analysis, 372, 277-379, 434
- Correlation coefficients, 374, 434
- Cortese, C. F., 327
- Costner, H. L., 374
- Cottrell, L. S., Jr., 260
- Council on Environmental Quality (CEQ), 318
- Cover letter, mail questionnaires, 149, 150, 435
- Cover story:
- in homosexual behavior experiment, 232
- in prophetic failure experiment, 226-227
- in qualitative research, 220-221
- Cox, B. G., 396
- Cox, F. L., 301
- Creighton, J., 318, 322, 333, 340
- Cronbach, L. J., 287, 290
- Cross-national studies, 263-264
- Cross-sectional study, 33, 442
- Cross-tabulation, 346, 435
- contingency tables and, 367-369
- Crowd behavior study, selective perception problems, 76
- Culture:
- cultural misperception, 12
- organizing principles and, 13
- Cumulative scaling, 386, 435
- Curiosity, types of, 29
- Data:
- Class I and Class IV data sets, 266-267
- errors in, 270-271
- observational, analysis of, 87-88
- primary data collection, 33
- problems of location and availability, 265-166
- secondary analysis, 33
- sources for secondary analysis, 268-269
- See also* Data analysis.
- Data analysis, 345-357
- coding, 346-356
- cross-tabulation and contingency tables, 367-369
- data entry, 356-357
- descriptive analysis, 358-367
- indexing and scaling, 384-387
- testing relationships between variables, 369-383
- Data collection:
- methods of, 33
- professional research organizations, 128-130
- See also* Survey research.
- Data entry, 346, 356-357, 435
- cleaning the data, 356-357
- Dat triangulation, 40
- Day, R. A., 397
- Day, R. C., 164-165, 168, 175, 179, 191, 242, 253-254
- Debriefing, 186-187
- behavior modification experiment, 194
- guaranteed income experiment, 199
- in selection for obedience to authority experiment, 189-190
- Definitions, operational, 45-46
- Degree of participation and secrecy, 219-220

- Degree of participation and secrecy (*cont.*)
 - in homosexual behavior experiment, 231
 - in prophetic failure experiments, 226
- Degrees of freedom calculation, 371
- Demographic impacts, 320
- Demographic items, 141, 144, 435
- Denissoff, R. S., 247
- Denzin, N. K., 14, 40, 108, 114, 115
- Dependent variables, 44 435
 - experimental research, 164
 - laboratory experiments, 166
 - measurement in experimental research, 175
- Description, in empirical science, 4
- Descriptive, 14
- Descriptive analysis, 346, 358–367, 435
 - frequency distribution, 358–362
 - frequency distributions and comparisons, 364–367
 - levels of measurement, 362–364
- Descriptive research, 33, 435
- Descriptive statistics, 369, 435
- Design, *See* Research design.
- Deutsch, M., 74, 79, 80, 86
- Deutscher, I., 211, 286, 303, 425
- Dickson, W. J., 11, 272
- Difference of means, 372, 435
- Difference of proportion, 372, 435
- Differential selection:
 - as an extraneous variable, 179–180
 - in static-group design, 183
- Dillman, D. A., 137–138, 146, 151
- DiRenzo, T., 393
- Discounting, 425, 435
- Disinterestedness, 8
- Disproportionate stratified sample, 61, 435
- Distribution questionnaires, 147–152
- Diversion program, *See* Juvenile diversion program evaluation.
- Divinney, L. C., 260
- Documentation, of data in archives, 266
- Dommel, W., Jr., 18
- Double-barreled questions, 117, 435
- Double rationale, 29
- Double screen, 425–427, 435
 - conceptual screen, 426–427
 - methodological screen, 426
- Draw-A-Man-Test, content analysis, 243
- DuBois, W. E. B., 21
- Duff, M. K., 316
- Dunphy, D. C., 247
- Dwelling units, used in sampling, 63

- Economic impacts, 320
- Economy, in writing, 403
- Edwards, A. L., 386
- Effort evaluation, 292, 436
- Einstein, A., 397, 401
- Elegance, 37, 436
 - in writing, 404
- Ellis, D., 93–95
- Ellsworth, P. C., 420
- Elms, A. C., 190, 423, 425
- Elpers, J. R., 21
- Empirical, 4, 436
- "Empirical assumption," 8
- Empiricism, principle of, 422
- Enke, C. G., 392
- Environmental impact assessment, 320, 436
- Environmental (social impact) research, 33
- Environmentalism, *See* Social Impact Assessment.

- Equality of Educational Opportunity* report, 163–160, 414–415
 - assumption of casual relationship, 168
 - as illustration of questionnaire survey, 153–160
- Eraut, M., 290
- Ethical issues, rights of human subjects, 15–20, 212–213, 218
- Ethnic origin, questionnaire format, 144–145
- Evaluation, types of, 283–284
- Evaluation research, 33, 282–310
 - design of, 290–296
 - juvenile diversion program evaluation, 304–309
 - methods of, 296–303
 - organizational context of, 285–290
- Evaluation Studies Review Annuals*, 283
- Evaluative research, 21
 - secondary analysis, 261
- Evans, J. W., 282, 288, 289
- Evolution, in writing, 404
- Exit:
 - in homosexual behavior experiment, 234
 - in prophetic failure experiment, 230
 - qualitative researcher, 223–224
- Expected frequencies, 370
- Experiment, 33, 436
- Experimental data, secondary analysis, Hawthorne experiments, 271–273
- Experimental design, 177–186
 - ABA Design, 185–186
 - behavior modification experiment, 194
 - extraneous variables, 177–180
 - guaranteed income experiment, 198–199
 - one-group pretest-posttest design, 181–182
 - one-shot case study, 180–181
 - pretest-posttest control group design, 183–185
 - program evaluation method, 300–301
 - in selection for obedience to authority experiment, 188–189
 - Solomon Four-Group design, 184–185
- Experimental mortality:
 - as an extraneous variable, 180, 436
 - in static-group design, 183
- Experimental research, 164–203
 - behavior modification experiment, 191–197
 - example experiments, 187–202
 - guaranteed income experiment, 197–202
 - obedience to authority experiment, 187–191
 - steps in, 174–187
 - strengths and weaknesses of, 167–170
 - subjects, 171–174
 - supervisor style and worker productivity, 164–165
 - types of experiments, 166–167
- Experimental research procedures, 174–187
 - analysis and report, 187
 - debriefing, 186–187
 - experimental design, 177–186
 - operational definitions development, 174–176
 - pilot study, 187
 - selecting subjects 176–177
 - statement of research problem, 174
- Experimental task, 175
- Experimenter bias, as an extraneous variable, 180
- Experiments, types of, 166–167
- Explained variance, 379, 436
- Exploratory interview, 104, 436
- External reality, 12
- External validity, 168–170, 436
- Extraneous factors, 186
- Extraneous variables, 177–180

- Face-to-face interviews, 137, 436
 - in social impact research, 334
- Factor analysis, 386, 436
- Federal agencies:
 - sources of data, 269, 330
 - survey restrictions, 332–333
- Ferritor, D., 93–95
- Festinger, L., 225–230
- Field experiments, 166–167, 436
- Field log, 82, 436
- Field setting, 28
- Fieldwork, reactive effects, 38
- Filstead, W. J., 33, 206, 207
- Fiscal impacts, 320
- Fixed-response items, 118
- Folk knowledge, 4
- Formal theory, 14, 436
- Franke, R. H., 272, 273, 422
- Freeman, H. E., 283, 285, 288, 294, 301
- Frequency distribution, 436
 - bimodal distribution, 362
 - comparison of distributions, 364, 367
 - cross tabulations, 367–369
 - percentage distributions, 359
 - skewed distribution, 361
 - standard normal distribution, 359–361
- Freudenburg, W. R., 316
- Gall, J., 296
- Gallimore, R., 290
- Gallup Poll, 128
- Gamma, 376–377
- Gans, H. J., 416–417
- Garfield, E., 393
- Gecas, V., 246
- Geer, B., 104, 116, 211, 335
- General Social Surveys (GSS), 129, 267–268
- Geographic clusters, 63
- Gerbner, G., 250–252
- Gianetto, R. H., 91
- Gilbert, J. P., 299, 300
- Gilmore, J. S., 316
- Glaser, B. G., 261–263
- Glasser, G. J., 55
- Glenn, N. D., 262, 267
- Glock, C., 218
- Goal displacement, 286
- Goal diversion, 286
- Goffman, E., 81
- Gold, M., 303
- Gold, R. L., 216
- Goldi, J., 128
- Goode, W. J., 77
- Gottfredson, M. R., 289
- Gould, J. D., 403
- Grand total, 368
- Green, R. L., 415
- Greer, S., 30
- Griffin, J. H., 219–221, 223–224
- Groeneveld, L. P., 197, 201
- Gross, L., 250–252
- Group settings, patterns of interaction category system (Bale's method), 84–85
- Groves, R. M., 55
- Guaranteed income experiment, 197–202
- Gubanich, A. A., 396
- "Guinea pig effect," 11, 38
- Guttentag, M., 301
- Guttman scales, 386
- Hagedoren, R., 426
- Hamblin, R. L., 93–95
- Handler, L., 243
- Hannan, M. T., 197, 201
- Hanson, J., 110
- Harding, T. S., 392
- Harris Research Organization, 128
- Harvey, C. D., 273, 275
- Hatt, P. K., 77
- Havens, A. E., 286
- Hawthorne experiments, 271–273, 421–422
 - Hawthorne effect, 11, 271, 437
- Head Start Program, 289–290
- Heisenberg principle, 11, 437
 - testing effect, 178
- Henle, R. J., 3
- Hennessey, M., 292
- Henshel, R. L., 419, 420, 421
- Henslin, J. M., 208, 209, 216, 217
- Hershey, N., 18
- Heterogeneity of population, 68
- Heyns, R. W., 83
- Hickman, E., 316
- Hingson, R., 138
- History, as an extraneous variable, 177–178, 437
- Hoinville, G., 22
- Holmes, C. C., 173
- Holmes, J. R., 173
- Holsti, O. R., 239, 250
- Holt, R. T., 33
- Homans, G. C., 14
- Homosexual behavior study, 231–235
- Hooper, D., 318, 322, 333, 340
- Hooper, D. A., 304, 305
- Hoover, K., 4, 7, 28, 43
- Horowitz, I. L., 410, 412
- Horton, P. J., 30
- Houts, K. C., 108, 110, 113, 246
- Howes, R. J., 243
- Hudson, C. R., 401, 404
- Hughes, E. C., 103
- Human faculties limitations, 12
- Human subjects, ethics of research with, 15–20, 212–213, 218
- Humphreys, L., 81, 212, 216, 231–235
- Hursh, G. D., 121, 123
- Hyman, H., 47, 260–261, 264, 268, 270
- Hypotheses, 14, 437
- Hypothesis, research, 44
- Hypothesis testing:
 - content analysis, 241
- Hypothesis-testing study, 33, 437
- Idle curiosity, 29
- Impact, 316, 437
- Impact analysis, in social impact assessment, 326–327
- Impact assessment, 284, 437
- Impact mitigation program, 340, 437
- Impact monitoring, 340–341, 437
- Impact projections, in social impact assessment, 325–326
- Impact statement, 22
- Impacts, types of, 320
- Implementation, of programs, 291–292
- Incentives, for research study participation, 148–150
- Independent variables, 44, 437
 - in experimental research, 164, 175–176
 - laboratory experiments, 166
 - tangible reinforcement, 193

- Index construction, 271
- Indexing, 346, 384, 437
- Indian activism study, 252-254
- Indian affairs relocation program, 275-276
- Indicators, 264
- Industry, social impact assessment and, 319-320
- Inferential statistics, 369, 437
- Information explosion, 392-393
- Informed consent, 19, 437
 - components of, 19-20
- Infrequent behavior, purposive sample, 65-66
- Instrument decay, as an extraneous variable, 178-179, 437
- Instruments, 37, 437
- Interaction effects, testing for, 372
- Internal validity, 177, 437
- Interuniversity Consortium for Political and Social Research, 266, 267, 268, 269
- Interval data:
 - measures of association, 377-383
 - tests of significance, 372
- Interval scale, 362-364, 437
- Interviewer effects, 38
- Interview guide, 105, 438
- Interview procedures, 120-126
 - approaching respondents, 124-126
 - selecting interviewers, 121-122
 - training interviewers, 122-124
- Interviews, 27, 437
 - program monitoring method, 297
- Interview schedule, 115-119
 - effective communication, 115-116
 - open and closed questions, 118-119
 - question formation problems, 116-118
- Interview studies, 33, 102-132, 437
 - conducting the interview, 120-126
 - data collection example, 128-130
 - interview, explanation of, 102-104
 - interview schedule, 115-119
 - interview schedule, pretesting, 120
 - issues in design of study, 105-115
 - reactivity in, 11-12
 - reliability of information study, 107-108
 - telephone interview, 126-128
 - types of interviews, 104-105
- Interview subjects:
 - approaching for the interview, 124-126
 - common questions of, 123-124, 125
 - See also* Issues in interview Studies.
- Investigator triangulation, 40, 438
- Involuntary error:
 - in interviews, 110-112
 - techniques for avoidance of, 114-115
- Issues in interview studies:
 - bias sources, 112-115
 - involuntary error problems, 111-112
 - psychology of respondent, 109-110
 - purposeful misinformation, 105-108
 - temporary role of respondent, 108-109
- Jacobs, H. A., 76
- Jacobs, J., 207, 208, 216
- Jahoda, M., 74, 79, 80, 86
- Janeczko, P., 396
- Janis, I. L., 260
- Johnson, A., 109, 138
- Johnson, T. W., 105
- Jones, B., 327
- Jordan, L. A., 128
- Journal Citation Reports*, 393
- Jowell, R., 22
- Juvenile diversion program evaluation, 304-309
- Kahn, R. L., 55, 103, 109, 112, 113, 114-115, 118, 119
- Kalton, G., 260
- Kaul, J. D., 272, 273, 422
- Kendall, P. L., 260
- Kenen, R., 81
- Kerner Commission, 413
- King, D. W., 393
- Kinsey, A. C., 109, 248, 250
- Kish, L., 65
- Klecka, W. R., 126
- Klein, M., 304
- Knowledge, 3-5
 - avenues to, 4
 - curiosity, 29
 - pure and applied research and, 27
 - social power and, 7
 - types of, 3-4
- Koenig, M., 393
- Kruzas, A. T., 269
- Kunz, P. R., 56, 140, 153
- Labeling, 89
- Laboratory experiments, 166, 438
- Labovitz, S., 426
- Lambda, 374-376, 438
- Lanham, R. A., 403
- Lantz, A. E., 316
- Lastrucci, C. L., 34, 426, 427, 428
- Lazarsfeld, P. F., 260, 261, 411, 421, 422
- Legislation, *See* Social Impact Assessment.
- Leistritz, F. L., 320
- Leslie, G. R., 30
- Lewis, O., 81
- Liebow, E., 395
- Liemohn, W., 19
- Light, R. L., 299, 300
- Likert, R., 385
- Likert scales, 385-386
- Lillydahl, J., 316
- Lindskold, S., 73
- Line of best fit, 378
- Literary knowledge, 4
- Longitudinal study, 33, 438
 - reversal of methods in secondary analysis, 263
- Love, L. L., 91
- Lovejoy, S. B., 316
- Lumsdaine, A. A., 260, 299
- Lumsdaine, M. H., 260
- Lundberg, G. A., 422
- Lynd, H. M., 242, 247, 366, 395
- Lynd, R. S., 366, 395
- Maccoby, E., 105
- Maccoby, N., 105
- Madsen, C. H., 192
- Magnetic tapes and disks, data entry, 356
- Mail questionnaires, 137, 147, 150-152, 438
- Malinowski, B., 211
- Mangione, T. W., 138
- Man-in-the-street sample, convenience sample, 65
- Manis, J. G., 30
- Marcus, A., 128
- Marginals, 358, 438
- Mark, M. M., 301
- Massey, M. E., 74
- Maturation:
 - as an extraneous variable, 178, 438

- in pretest-posttest design, 182
- McDill, E. L., 289, 290
- McDill, M. S., 289, 290
- McDonald, D. D., 393
- McKeown, R. L., 316
- McLean, J., 269
- McNemar, Q., 171
- McPhee, W. N., 422
- Mean, 358, 362, 438
 - standard normal distribution, 359
- Measurement, 42–43, 438
 - as change agent effect, 11, 38
 - levels in descriptive analysis, 362–364
- Measures of association, 369, 372–383, 438
 - Gamma, 376–377
 - Lambda, 374–376
 - multiple correlation and regression analysis, 379–382
 - partial correlation and beta coefficients, 382–383
 - simple correlation and regression, 377–379
- Measures of central tendency, 358
- Mechanical recording devices, 82, 438
- Median, 358, 362, 438
 - standard normal distribution, 359
- Mendenhall, W., 53
- Merton, R., 8, 14, 28, 29, 261, 420–421
- Messages, content analysis of, 240, 244
- Method of equal-appearing intervals, 384–385, 438
- Method of summated ratings, 385–386, 438
- Methodological approaches, 41–42
- Methodological triangulation, 40, 438
- Metz, W. C., 319
- Metzgar, G. D., 55
- Michigan Survey Research Center, 128
- Milgram, S., 18, 175, 187–191
- Miller, R. D., 18
- Mills, C. W., 22, 395, 425
- Misperception, in observation, 12–13
- Mode, 358, 362, 438
 - standard normal distribution, 359
- Moen, E., 316
- Moffitt, R. A., 197, 201
- Morgan, M., 250–252
- Moser, C. A., 260
- Mosteller, F., 299, 300
- Moynihan, P., 201
- Mueller, H. H., 374
- Multimethod approaches, 41, 438
- Multiple correlation and regression analysis, 379–382
- Multiple indicators, 47, 438
- Multiple operationalism, 11, 438
- Multiple regression analyses, 156, 272, 380, 439
- Multistage cluster sample, 62
- Murdock, S. H., 320, 322, 323, 332, 337
- Murphy, F. J., 46

- Namboodiri, N. K., 65
- Naroll, R., 259
- National Commission on Protection of Human Subjects of Biomedical and Behavioral Research, 19
- National Environmental Policy Act (NEPA), 317–318, 319, 340
- National Opinion Research Center (NORC), 128–130
 - data collection, 268
- National policy making and social research:
 - civil disorders report, 413
 - equality of educational opportunity study, 414–415
 - planning and decision making, 415–416
 - violence in America reports, 413–414
- National Research Act, 19

- Natural experiments, 167, 439
- Natural setting:
 - qualitative research, 211
 - unstructured observations, 79
- Nawy, H., 74
- Nay, J. E., 303
- Nea-saying, 38
- Negotiation, evaluation research and, 287–288
- Nelson, H., 105
- Nett, R., 8
- Nettler, G., 298
- Neubeck, K. J., 197, 199
- Newcomb, T., 2
- Nominal data:
 - Chi-square, 370–371
 - measure of association for, 374–376
- Nominal measurement, 436, 439
- Nonrandom sampling, 65–67, 439
 - convenience sample, 65
 - purposive sample, 65–66
 - quota sample, 66–67
 - snowball sample, 66
- Nonreactive research, 34
- "Notes on Problem-Finding in Sociology," (Merton), 28
- Null hypothesis, 370, 439

- Obedience to authority experiment, 187–191
- Observation, 11, 73–74
 - limits of scientific, 10–14
 - program monitoring method, 297
- Observational data, analysis of, 87–88
- Observational methods, 73–97
 - advantages of, 74–76
 - analyzing data, 87–88
 - examples of use, 88–96
 - problems of, 76–77
 - research, 78–87
- Observational research, 78–87
 - category systems, 83–85
 - content of observation, 81–82
 - recording observation, 82–83
 - sampling behaviors for, 85–86
 - setting structures, 79–81
 - training observers, 86–87
- Observed frequencies, 370
- "Observer" assumptions, 8
- Obtrusive observation, recording devices, 82
- "Obviousness" charge, about social science research, 420–423
- Occam's razor, 37
- Occupational data, 354–355
- Official goals, 286, 439
- Ogilvie, D. M., 247
- Oksenberg, L., 109, 110, 111–112
- OMB Circular A-40, 332–333, 439
- "On Intellectual Craftsmanship" (Mills), 22
- One-group pretest-posttest design, 181–182
- One-shot case study, 180–181
- Open-ended item, 104, 118–119, 439
 - questionnaires, 141–142
- Operational definitions, 45–46, 439
 - behavior modification experiment, 192–193
 - content analysis, 248–249
 - development in experimental research, 174–176
 - guaranteed income experiment, 197–198
 - in Indian activism study, 253–254
 - obedience to authority experiment, 187–188
 - reliability, 46–47
 - validity, 47–48

- Operational definitions (*cont.*)
 - in violence profile study, 251-252
- Operationalization, 264
- Operative concepts, 14
- Operative goals, 286, 439
- Ordinal data:
 - Chi-square, 370-371
 - measure of association for, 376-377
- Ordinal measurement, 326-336, 439
- Organization, of a research report, 399-401
- Organizational factors, evaluation research, 285-290
 - negotiation and, 287-288
 - organizational goals, 285-287
 - political interests and, 288-290
- Organizations, *See also* Evaluation research.
- Organized skepticism, 8
- Originating questions, 28
- Orwell, G., 402
- Ott, L., 53

- Palm, R., 316
- Palmer, A. M., 269
- Panek, P. E., 174
- Parameters, 53, 370, 439
- Park, R., 21
- Parsimony, 37-38, 440
- Partial correlation, 382-383, 440
 - beta coefficients and, 382-383
- Participant observation, 27, 33, 440
 - cab driver study, 208-210
 - degree of participation, 216-217
 - degree of secrecy, 217-218
 - in homosexual behavior experiment, 232-234
 - legal, moral, and injury risks, 213-214
 - limitations, 76-77
 - over identification with subjects, 214-215
 - in prophetic failure experiment, 227-230
 - relationship to observed, 80-81
 - requirements of, 208
 - researchers impact on group, 212-213
 - in social impact assessment, 335-337
 - unstructured data collection, 214
 - unstructured observations, 79
- Paskar, J., 269
- Pattison, M., 316
- Patton, M. Q., 288, 291-292, 293, 303
- Pattullo, E. L., 20
- Percentage distributions, 359
- Perry, J. B., Jr., 55
- Person triangulation, 40
- Personal delivery technique, 136, 440
- Personal documents, content analysis source, 245
- Pettigrew, T. F., 415
- Phillips, J. S., 174
- Pilot study, 187
 - behavior modification experiment, 195-196
 - guaranteed income experiment, 199
 - in selection for obedience to authority experiment, 190
- Policy problems, 30
- Political interests, evaluation research and, 288-290
- Political polling, 102
- Polk Company, 54
- Pooling, of multiple surveys, 270
- Population, 53, 64, 440
 - questionnaire surveys, 141
 - standard deviation of, 68
- Population explosion, 30
- Population study, *See* Sampling.
- Positional approach, 334

- Postcard follow-up, to mail questionnaires, 151
- Practical curiosity, 29
- Precoded, 347, 440
- Prediction, accuracy of social research, 419-420
- Presser, S., 116, 119
- Pretest-posttest control group design, 183-184
- Pretesting (interview schedule), 120, 440
 - questionnaires, 167
- Primary data collection, 33, 440
- Primary objectives, of research projects, 33
- Primary sources, 259-260
- Principle of empiricism, 422, 440
- Printed media, content analysis source, 245
- Pritchard, J., 128
- Probability, 370, 440
- Probe questions, 118, 130, 440
 - questionnaires and, 138
- Problems:
 - in social research, 30
 - scientific, 29-30
- Process evaluation, purpose of, 292
- Program evaluation methods, 298-303
 - design, 294-296
 - experimental design, 300-301
 - qualitative methods, 302-303
 - quasi-experimental design, 301-302
- Program monitoring, 284, 440
 - design of, 291-294
 - methods for, 297
- Program planning, 283-284, 440
- Project characteristics, 322, 323
- Project description, in social impact assessment, 325
- Projective personality tests, content analysis, 242, 440
- Prophetic failure experiment, 225-230
- Proportional reduction in error (PRE), 374, 440
- Propositions, 14, 440
- Pruning, writing process and, 403, 404
- Pure research, 27, 441
- Purposeful misinformation, in interview study, 105-106
- Purposive (judgment) sample, 65-66, 441

- Qualitative measurement, 33
- Qualitative research, 206-236, 441
 - cab driver study, 208-210
 - conducting of, 218-224
 - design, 215-218
 - homosexual behavior study, 231-235
 - problems, 212-215
 - program evaluation method, 302-303
 - prophetic failure study, 225-230
 - quantitative methods used in study, 210-211
 - strengths, 211-212
- Qualitative research design, 215-218
 - degree of participation, 216-217, 219-220
 - degree of secrecy, 217-218, 219-220
- Qualitative research procedures, 218-224
 - analysis of data and report writing, 224
 - cover and exit stories, 220-221
 - degree of participation and secrecy, 219-220
 - exit, 223-224
 - participation, 221
 - topic selection, 218-219
- Quantitative analysis, content analysis, 241-242
- Quantitative measurement, 33
- Quantitative method, qualitative researchers' rejection of, 207
- Quantitative-qualitative dimension, of research project, 32
- Quasi-experimental design, program evaluation method, 302-303

- Questionnaire surveys, 135–161
 - advantages of, 137–138
 - conducting, 139–153
 - disadvantages of, 138–139
 - example of (Equality of Educational Opportunity), 153–160
 - reactivity in, 11–12
 - religiosity and delinquency study, 135–137
- Questionnaire surveys procedures, 139–153
 - administering the questionnaire, 147–152
 - analysis and report, 152–153
 - population and sample selection, 141
 - questionnaire construction, 141–147
 - statement of a problem, 139–141
- Questionnaires, 27, 135, 441
 - in social impact research, 334–335
- questions for, 141–145
- Questions:
 - in interview studies, 116–119
 - questionnaire construction, 141–145
- Quota sample, 66–67, 441
- Random sampling, 53–65, 441
 - area sample, 62–65
 - cluster sample, 61–62
 - disproportionate stratified sample, 61
 - restricted sample, 57–59
 - sampling frames, 54–57
 - stratified sample, 59–61
 - systematic sample, 59
 - unrestricted sample, 57
- Range, 365, 441
- Rapport, 106, 441
- Ratio scales, 364, 441
- Rational, 28–29
 - double, 29
- Ravetz, J. R., 22, 29
- Reactive effects, 11, 77, 441
 - fieldwork and, 38–39
- Recording observation, aids in, 82–83
- Records, use in program monitoring method, 297–298
- Reed, J. S., 270
- Reeder, L., 128
- Rees, A., 197
- Regression, 272
- Regression analysis, 156, 377–379, 441
 - multiple regression analysis, 156, 272, 380
- Regression effect, 179
- Reliability, 40, 46–47, 441
 - in content analysis, 249–250
- Remus, W., 396
- Replication, 264
 - internal, external, intrasurvey, 270
- Report of the National Advisory Commission on Civil Disorders*, 413
- Reports, use in program monitoring method, 297–298
- Reputational approach, 334
- Research:
 - characteristics of, 9–10
 - definition of, 9
 - with humans, ethics of, 15–20
 - proper procedures, 9
 - purposes of, 20–22
 - theory and, 14–15
- Research design, 26–49, 441
 - principles of, 37–42
 - research languages, 42–48
 - selection of topics, 28–32
 - stages in, 34–37
 - types of, 32–34
- Research design, principles of, 37–42
 - allocation of resources, 39–40
 - methodological issues, 41–42
 - parsimony, 37–38
 - triangulation, 40–41
 - unobtrusiveness, 38–39
- Research hypothesis, 44
- Research instrument, effect of changes in, 38
- Research methods:
 - analyzing data, 345–388
 - appropriateness of, 426
 - content analysis, 239–256
 - evaluation research, 282–310
 - experimental research, 164–203
 - observational methods, 73–97
 - qualitative research, 206–236
 - sampling, 52–70
 - secondary analysis, 259–278, 442
 - Social Impact Assessment, 314–342
 - survey research, interview studies, 100–132
 - survey research, questionnaire surveys, 135–161
- Research project:
 - implementation, 37
 - justification of, 36
 - methods justification, 36–37
 - specifying objectives, 35–36
 - stages of, 34–37
 - utilization of findings, 36
- Research proposal, 33, 441
- Research terminology, 42–48
 - measurement, 42–43
 - operational definition, 45–46
 - reliability, 46–47
 - research hypothesis, 44
 - validity, 47–48
 - variables, 43–44
- Researchers, independent, 262–263
- Resource allocation, research design principle, 39–40
- Resource efficiency, 284, 441
- Response hierarchy, of verbal comments, 242
- Response rate, questionnaires, 152, 441
- Response sets, 11, 38
- Restricted sample, 57–59, 441
- Retrospective, 33, 442
- Retrospective data, 101
- Reznikoff, M., 243
- Rice, S. A., 112
- Riecken, H. W., 225–230
- Ritualism, 28, 442
- Roach, J. L., 197, 199
- Roderer, N. K., 393
- Roethlisberger, F. J., 11, 272
- Roland, C. G., 396
- Role selection effect, 11, 38
- Roper Public Opinion Research Center, 128, 261
 - data distribution, 268
- Rorschach test, 242
- Rosenhan, D. L., 89–91
- Rosenthal, R., 171, 173, 174, 180
- Rosnow, R. L., 171, 173, 174
- Rossi, P. H., 21, 22, 282–283, 285, 288, 290, 293, 294, 296, 299, 301
- Row totals, 368
- Rubenstein, C., 173
- Rubin, Z., 170
- Rubinstein, E. A., 239, 241
- Rush, M. C., 174
- Ryan, L., 397
- Sachs, H. L., 399

- Sample, 52, 442
 - experimental subjects, 173
 - sample size, 67-69
- Sampling, 47, 52-70
 - behaviors for observation, 85-86
 - methods, 37
 - nonrandom, 65-67
 - questionnaire surveys, 141
 - random, 53-65
 - sample size, 67-69
 - telephone interviews, 126
- Sampling frame, 53, 54-57, 442
- Sampling unit, 53, 442
- Samuel, W., 191
- Sanders, W., 241
- Scaling, 346, 384-387, 442
 - cumulative scaling, 386
 - method of equal-appearing intervals, 384-385
 - method of summated ratings, 385-386
- Scalogram analysis, 386
- Scanlon, J. W., 303
- Schachter, S., 225-230
- Schedules, 82, 442
- Schmidt, R. E., 303
- Schuessler, K. F., 374
- Schuman, H., 106, 113, 116, 119
- Schwartz, H., 207, 216
- Science, 3, 442
 - elements of, 7-8
- Science Citation Index*, 392-393
- Scientific disciplines, problems intrinsic to, 30
- Scientific knowledge, 4
- Scientific method, 4, 6-10, 442
 - "articles of faith," 8
 - assumptions of, 8-9
 - basic values, 8
 - research characteristics, 9-10
- Scientific observation, 75-76
 - limits of, 10-14
 - misinterpretation in communication of interpretations, 13-14
 - misperception and, 12
 - misrepresentation in interpretation, 13
 - problems of, 76-77
 - reactive effects, 10-12
- Scientific problems, 29-30
- Scientific research project, stages of, 34-35
- Scott, W. R., 295
- Screening questions, 142, 442
 - examples of, 143
- Sechrest, L., 299, 300
- Secondary analysis, 33, 259-278, 442
 - advantages of, 261-264
 - conducting secondary analysis, 267-271
 - designs for, 270
 - disadvantages of, 265-267
 - of evaluative research, Indian relocation study, 275-276
 - of experimental data, Hawthorn experiments, 271-273
 - of survey data, widowhood status study, 273-275
 - secondary data search, 268-269
- Secondary data:
 - sources for community impact analysis, 329
 - analysis in social impact assessment, 327-337
- Secondary sources, 259-260
- Secrecy, in participant observation, 217-218, 219-221
- Selective perception, 76, 442
- Self-report, 46
- Selltiz, C., 74, 79, 80, 85, 86
- Sermons, content analysis source, 247
- Sesame Street* evaluation, 299
- Setting factors:
 - natural setting, 79, 211
 - observation of, 81
- Sewell, W. H., 411
- Sheaffer, R. L., 53
- Sherrow, F., 373
- Shirley, M. M., 46
- Shortland, R., 175
- Shrader, R. R., 243
- Sibley, E., 410
- Significance, 28
- Significance level, 370, 442
- Signorielli, N., 250-252
- Silverman, L. T., 239
- Simon, J. L., 260
- Simple correlation, 378-379
- Simple linear regression, 377-378, 442
- Site area characteristics, 321-322, 323
- Sjoberg, G., 8
- Skewed distribution, 361, 442
- Small effects, 299
- Small group setting, structured observations, 80
- Smith, M. B., 260
- Smith, M. S., 247
- Smith, P., 304, 305, 307
- Smith, R. J., 19
- Snowballing techniques, 66, 334, 442
- Social action programs:
 - goals, 286-287
 - negative evaluation, 289
 - unexpected/unintended effects, 295-296
- Social change, content analysis, 241
- Social Impact Assessment (SIA), 314-342, 442
 - conducting social impact assessment, 322-339
 - elements of, 320-322
 - impact mitigation and monitoring, 339-341
 - purpose of, 316-320
- Social impact assessments procedures, 322-339
 - task definition, 324-327
 - triangulated approach, 327-339
- Social impact research, 33, 442
- Social impacts, 320
- Social learning theory, ABA design, 185-186
- Social philosophy, problems in, 30
- Social problems:
 - dangers in study of, 31
 - as research topics, 30-31
- Social programs, implementation, 291-292
- Social science:
 - purpose of, 3
 - defense of writings in, 393-397
- Social Science Citation Index*, 393
- Social science data, collection of methods approach, 11-12
- Social science research:
 - applications of, 411-412
 - controversial nature of, 418-419
 - experimental subjects and, 171, 173
 - individual consumer and, 423-429
 - national policy making and, 412-416
 - purposes of, 26
 - utilization of, 409-429
 - utilization of, barriers to, 416-423
- Society, policy problems, 30
- Sociological method, interviewing, 103
- Sociology, problem-finding in, 28-29
- Software, 355, 443
- Solomon, M. A., 283

- Solomon-Four Group design, 184-185
- Sorensen, J. E., 21
- Source material, primary and secondary, 259-260
- Space triangulation, 40
- Specifying questions, 29
- Sponsors, of research studies, 148
- Spradley, J. P., 216
- Sprafkin, J. N., 239
- Sprehe, J. T., 289-290
- Standard deviations, 68, 383
- Standard normal distribution, 359-361, 443
- Stanley, J. C., 170, 177, 179, 180-181
- Star, S. A., 260
- Statement of research problems, 174
- behavior modification experiment, 191-192
- content analysis, 248
- guaranteed income experiment, 197
- in Indian activism study, 252-253
- obedience to authority experiment, 187
- in violence profile study, 250
- Static-group comparison design, 182-183
- Statistic, explanation of, 53
- Statistical regression, as an extraneous variable, 179
- Statistical significance, 364, 369, 443
- Statistics:
- multiple regression analyses, 156
- regression analysis, 156
- regression effect, 179
- testing relationships between variables, 369-383
- See also* Data analysis.
- Stauss, J. M., 56, 160
- Steady state, 185
- Stone, P. J., 247
- Stouffer, A. S., 260
- Strata, 59
- Stratified sample, 59-61, 443
- Street Corner Society* (Whyte), 77
- Strommen, M., 109, 125, 138
- Structure in research design, 79-81
- Structured interviews, 104
- Structured observations, 80, 443
- classroom behavior change study, 93-96
- recording devices, 82-83
- smiles in parental communication study, 91-92
- Strunk, W., Jr., 402
- Stufflebeam, D. L., 287, 290
- Style, of a research report, 401-402
- Subjects:
- in behavior modification experiment, 193-194
- debriefing, 186-187
- for experimental research, 171, 173-174, 176-177
- as an extraneous variable, 179-180
- in guaranteed income experiment, 198
- in obedience to authority experiment, 188
- Substantive significance, 369
- Suchman, E. A., 260, 287, 290, 294
- Sudman, S., 55, 65, 137, 144
- Survey methods, in social impact assessment, 331-335
- Survey research, 100-102
- advantages of, 101-102
- disadvantages of, 102
- interview studies, 102-132
- questionnaire surveys, 135-161
- sampling, 53
- Surveys, use in program monitoring method, 298
- Swift, J., 401
- Systematic observation, 7, 443
- Systematic sample, 59, 443
- Table of Random Numbers, 58
- Takata, G., 292
- Tangible reinforcement, 193
- Tasks in social impact assessment:
- baseline profile, 324
- baseline projections, 324-325
- impact analysis, 326-327
- impact projections, 325-326
- project description, 325
- Tax rolls, as sampling frames, 55
- Tearoom Trade* (Humphreys), 234-235
- See also* Homosexual behavior study.
- Technology, 7, 443
- Tedeschi, J. T., 73
- Telephone directories, as sampling frames, 55, 59
- Telephone interviews, 126-128, 137, 443
- advantages of, 126-127
- disadvantages of, 127-128
- in social impact research, 334
- Television programs, content analysis, 239-241, 246-247
- See also* Violence Profile Study.
- Testing, in empirical science, 4
- Testing effect, 178 443
- as an extraneous variable, 178
- in pretest-posttest design, 182
- Testing relationships between variables, 346, 369-383
- measures of association, 372-383
- tests of significance, 370-372
- Tests of significance, 369, 370-372, 443
- analysis of variance, 372
- Chi-square, 370-371
- correlation analysis, 372
- difference of means test, 372
- difference of proportions test, 372
- Tharp, R. G., 290
- Thematic Apperception Test, 242
- Theory, 10, 443
- characteristics of, 14
- Theory triangulation, 40
- Thomas, D. L., 31
- Thomas, D. R., 192
- Thomas, J. K., 320, 322, 323, 332, 337
- Thomas, W. I., 245
- Thompson, J., 318, 322, 333, 340
- Thurstone, L. L., 384
- Thurstone scales, 384-385
- Tilly, C., 414
- Time-orientation, of research projects, 33
- Time triangulation, 40
- Tompkins, T., 304, 307
- Topic selection: 28-32
- homosexual behavior experiment, 231
- prophetic failure experiments, 225
- qualitative research, 218-219
- significance assessment, 28-30
- social problems, 30-31
- topics of double relevance, 29
- Training programs:
- coder training, 249
- interviewers, 121
- observation, 86-87
- Treatment evaluation, 293, 443
- Tremblay, K. R., 128
- Trends, content analysis, 245
- Triangulated approach, social impact assessment, 327-337
- importance of, 337-339
- participant observation, 335-337
- secondary data sources and analysis, 327-331
- survey methods, 331-335

- Triangulation, 40–41, 443
 accuracy and, 77
 Trochim, W. M. K., 268, 269
 Trow, M., 41–42
 Trumble, R. R., 18
 Truncation, of multiple surveys, 270
 Tuchfarber, A. J., 126
 Tuma, N. B., 197, 201
 Turner, J. E., 33
 Tuskegee Syphilis Study, 16–18
 Twenty Statement Test, 242
 Two-by-two table, 367–368
- Underwager, R., 109, 125, 138
 Unidimensionality, 386, 443
 Unit of analysis, 248–249, 443
 Universalism, 8
 Universe, 53, 64, 443
 Unobtrusive data, 39, 443
 Unobtrusive observation, recording devices, 82
 Unobtrusive research, 33, 443
Unobtrusive Measures: Nonreactive research in the Social Sciences (Webb et al.), 11, 39, 41
 Unrestricted sample, 57, 443
 Unstructured observation: 79–80, 443
 psychiatric hospitals study, 89–91
 use of field log, 82
 Utility company records, as sampling frames, 55
- Validity, 14, 40, 47–48, 443
 Van Nostrand, A. D., 398
 Variables, 29, 43, 443
 testing relationships between, 369–383
 Verbal comments, content analysis, 242
 Verbal communication, observation of, 82
 Verification, 6, 443
 Vidich, A., 105, 109, 110
Violence in America, 413–414
 Violence Profile Study, 250–252
 Volunteers, experimental subjects, 173–174
- Wales, E., 248
 Wallenstein, B., 397
 Warner, W. K., 286
 Wasserman, P., 269
 Watson, K. S., 320
 Wattenberg, B., 30
 Watts, H. W., 197
 Wax, R. H., 220
 Webb, E., 11, 38, 39
 Webster, W. J., 287, 290
 Weick, K. E., 420
 Weigert, A. J., 423
 Weiss, C. H., 283, 287, 288, 410, 411–412, 417, 423
 Weisz, R., 316
When Prophecy Fails, 230
 See also Prophetic failure experiment.
 White, E. B., 402
 Whitehead, A. N., 424
 Wholey, J., 303
 Whyte, W. F., 77, 212, 214
 Widowhood status study, 273–275
 Wilderness area study, 314–315, 415–416
 Wilensky, H. L., 411
 Williams, J. A., 106
 Williams, R. M., Jr., 260, 397
 Williams, W., 21, 22, 282, 285, 288, 289, 290
 Williamson, S. T., 395–396, 399
 Winick, C., 74–75
 Wiseman, F., 137, 139
- Within-method triangulation, 40–41, 443
 Witmer, H. L., 46
 Word-for-word responses, computer analysis, 355–356
 Wortman, P. M., 261, 265, 268
 Wright, J. D., 282–283, 293
 Wright, S. R., 282–283, 285, 288, 293, 294, 299, 307
 Writing about research, 391–405
 audience identification, 39
 basic principles, 404
 experimental research, 187
 organization, 399–401
 scientific creativity of, 397–399
 scientific writing, 392
 social science writing, defense, 393–397
 sociological classics, 395
 style, 401–402
 workshops and guidebooks, 396–397
 writing process, 402–404
- Yea-saying, 38, 110
 Yeaton, W. H., 299, 300
- Zander, A. F., 83
 Zelditch, M., Jr., 420
 Zeller, M. E., 319
 Ziman, J. M., 7, 8
 Zinsser, W., 397
 Znaniecki, F., 245

ISBN 0-13-818336-8